INSTITUTO MILITAR DE ENGENHARIA

WILLIAM AUGUSTO RODRIGUES DE SOUZA

IDENTIFICAÇÃO DE PADRÕES EM CRIPTOGRAMAS USANDO TÉCNICAS DE CLASSIFICAÇÃO DE TEXTOS

Dissertação de Mestrado apresentado ao Curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Sistemas e Computação.

Orientador: José Antonio Moreira Xexéo - D. Sc.
Co-orientadora: Cláudia Maria G. M. de Oliveira - Ph.D.

Rio de Janeiro
2007

INSTITUTO MILITAR DE ENGENHARIA
Praça General Tibúrcio, 80 – Praia Vermelha
Rio de Janeiro – RJ CEP 22290-270

S729i Souza, William Augusto Rodrigues de.
Identificação de Padrões em Criptogramas usando Técnicas de Classificação de Textos / William Augusto Rodrigues de Souza – Rio de Janeiro: Instituto Militar de Engenharia, 2007.
252 p.: il., graf., tab.

Dissertação (mestrado) – Instituto Militar de Engenharia, 2007

1. Identificação de padrões. 2. Criptoanálise. 3. Criptografia. 4. Recuperação de Informação. 5. Mapa de Kohonen
I. Título II. Instituto Militar de Engenharia.
CDD 003.54

INSTITUTO MILITAR DE ENGENHARIA

WILLIAM AUGUSTO RODRIGUES DE SOUZA

IDENTIFICAÇÃO DE PADRÕES EM CRIPTOGRAMAS USANDO TÉCNICAS DE CLASSIFICAÇÃO DE TEXTOS

Dissertação de Mestrado apresentada ao Curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Mestre em Sistemas e Computação.

Orientador: José Antonio Moreira Xexéo - D. Sc.

Co-orientadora: Cláudia Maria G. M. de Oliveira - Ph.D.

Aprovada em 27 de fevereiro de 2007 pela seguinte Banca Examinadora:

_____

Prof. José Antônio Moreira Xexéo – D.Sc. do IME – Presidente

_____

Profa. Cláudia Maria G. M. de Oliveira - Ph.D. do IME

_____

Prof. Geraldo Bonorino Xexéo - D.Sc. da COPPE/UFRJ

_____

Prof. Luís Alfredo Vidal de Carvalho - D.Sc. da COPPE/UFRJ

Rio de Janeiro
2007

À Maria Izabel: minha mãe, “meu pai”, minha amiga.

AGRADECIMENTOS

À minha família, em especial à Simone, pelo amor e paciência.

Ao Professor José Antônio Moreira Xexéo, pela orientação precisa e objetiva, pela confiança e pelas oportunidades oferecidas tanto na área da pesquisa como na de docência.

À Professora Cláudia Maria Garcia Medeiros de Oliveira, pela orientação, apoio e atenção ao longo de todo o curso.

Ao Eduardo Zapico Mouro, pelo incentivo e ajuda constante na minha busca pelo conhecimento.

Ao Paulo Sérgio Pagliusi pelo valoroso aconselhamento.

À Olya Guennadievna Khamyanova, à Ellyn Dinkelmann Girdwood e ao Leandro de Araújo Loures Coelho pela contribuição a este trabalho.

À Marinha do Brasil e ao Exército Brasileiro pela oportunidade.

Ao Centro de Análise de Sistemas Navais pela confiança depositada na minha pessoa para representá-lo.

À Escola de Guerra Naval pelo inestimável apoio prestado durante todo o curso

Ao Instituto Militar de Engenharia e aos integrantes (professores, alunos e funcionários) da Seção de Engenharia de Computação pela acolhida, pela convivência harmoniosa e pelo ambiente adequado ao desenvolvimento de pesquisas.

A todos aqueles que contribuíram direta ou indiretamentamente para o sucesso deste trabalho.

“Se a necessidade é a mãe das invenções, então a adversidade é a mãe da criptoanálise.”

SIMON SINGH

SUMÁRIO

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE GRÁFICOS

LISTA DE FÓRMULAS

LISTA DE ABREVIATURAS E SÍMBOLOS

ABREVIATURAS

AES	–	Advanced Encryption Standard.
DES	–	Data Encryption Standard.
NBS	–	National Bureau of Standards.
NIST	–	National Institute of Standard and Technology.
NESSIE	–	New European Schemes for Signatures, Integrity and Encription.
RSA	–	Rivest, Shamir e Adleman

SÍMBOLOS

$\oplus$	–	operação lógica "OU EXCLUSIVO" ou *XOR* ou soma modulo dois.
$\otimes$	–	multiplicação modular.

RESUMO

O desenvolvimento de produtos criptográficos requer meios para se avaliar a confiabilidade destes produtos. Desta forma, pode ser definido um conjunto de requisitos a serem atendidos para que um algoritmo criptográfico seja considerado confiável. Um desses requisitos poderia ser a inexistência de padrões nos criptogramas gerados por esses algoritmos.

No trabalho de Carvalho (2006), foram detectados padrões em criptogramas gerados com os algoritmos DES e AES, com chaves de 64 e 128 *bits*, respectivamente, no modo de operação ECB. Estes padrões se caracterizaram pelo agrupamento dos criptogramas cifrados com a mesma chave.

Neste contexto, a presente dissertação tem o objetivo de identificar padrões em criptogramas gerados pelos algoritmos DES, com chaves de 64 *bits*, AES, com chaves de 128, 192 e 256 *bits*, e RSA, com chaves de 64, 128, 256, 512 e 1024 *bits*, por meio de técnicas de Recuperação de Informações e de Inteligência Artificial. Além disso, o modo de operação CBC é avaliado.

ABSTRACT

The development of cryptographic products to require ways to evaluate the reliability of these products. In this way, a set of requirements can be defined in order to the cryptographic algorithm to be considered reliable, since those requirements have been accomplished. One of these requirements could be the inexistence of standards in the cryptograms generated for those algorithms.

In the work of Carvalho (2006), some standards was found in cryptograms generated with algorithms DES and AES, with keys of 64 and 128 bits, respectively, using the mode ECB. These standards were characterized by means of cryptogram clustering that was ciphered with the same key.

In this context, the present work has the objective to identify standarts in cryptograms generated with algorithms DES, with keys of 64 *bits*, AES, with keys of 128, 192 and 256 *bits*, and RSA, with keys of 64, 128, 256, 512 and 1024 *bits*, by means of Information Retrieval and Artificial Intelligence techniques. Moreover, the mode CBC is evaluated.

1 INTRODUÇÃO

O processo de criptografia c s funções são descritas por algoritmos que são co as funções tem a finalidade de criar um texto ilegí ção é utilizada para voltar ao texto legível a partir unções podem ser chamadas de função de cifrar e pectivamente. Aos textos ilegíveis podemos cham emos chamar de texto em claro.

Tanto a função de cifrar como a algoritmos, os quais descrevem os métodos utili ura 1.1).

da chave utilizada na função de cifrar. Alternativamente, pode ter o objetivo de descobrir a própria chave utilizada nessa função.

Podemos utilizar como suporte à atividade de criptoanálise alguns indicadores de fraqueza no processo de cifrar, para que tenhamos uma direção a seguir na criptoanálise, como , por exemplo, a existência de padrões nos criptogramas. Assim, observando o passado, vemos que a criptoanálise clássica era fortemente baseada nas características lingüísticas do idioma de origem do texto em claro (SINGH, 2003). Desta maneira, a criptoanálise explorava as propriedades intrínsecas do idioma refletidas nos criptogramas. Dentre estas, podemos citar: freqüências de letras, freqüência de palavras e ocorrência de n-gramas[1] comuns em palavras. Com a evolução do poder computacional e o surgimento de técnicas modernas de criptografia, que exploravam este novo poder, as técnicas de criptoanálise baseadas em características lingüísticas tornaram-se cada vez mais escassas na literatura.

Entretanto, enquanto as técnicas de criptografia evoluíam, outro ramo da ciência da computação, a lingüística computacional, também se desenvolveu de forma expressiva, produzindo métodos e técnicas eficientes para tratar as características lingüísticas aplicadas à computação (MITKOV, 2005). Esses métodos e técnicas, em especial a classificação de informações textuais, podem se constituir em matéria de especial interesse para a criptoanálise, isto porque, embora as técnicas atuais de criptografia envolvam problemas matemáticos difíceis (MENEZES, 1996) e (NIST, 2001), a sua matéria-prima continua sendo a mesma: o texto[2].

1.1 MOTIVAÇÃO

O estado da arte em lingüística computacional é suficiente para motivar uma investigação do uso de suas técnicas aplicadas a criptoanálise. O que foi observado, até o presente momento, é que tal uso permanece quase inexplorado, o que também é outro aspecto motivador da investigação.

Para ilustrar esta situação, podemos citar o exemplo da conferência EUROCRYPT, realizada anualmente pela Associação Internacional de Pesquisa em Criptologia. Foram consultados 294 artigos, relativos às conferências de 1998 a 2004 e 2006, dos quais 51

[1] Conjunto de 1, 2, 3 ou n letras.
[2] Outros tipos de dados podem ser cifrados, como: sons, imagens e vídeos (SCHNEIER, 1996). Mas o nosso interesse no presente trabalho são os textos apenas.

tratavam do tema criptoanálise e três tratavam do tema recuperação de informações privadas em banco de dados. Nos textos analisados não foram encontradas quaisquer referências ao assunto apresentado neste trabalho. Os três textos sobre recuperação de informações privadas em banco de dados se constituem em assunto diverso ao normalmente tratado pela recuperação de informações no contexto da lingüística computacional. Para maiores detalhes sobre os artigos analisados e seus respectivos autores, consultar o apêndice 1.

Uma vez que a classificação de informações textuais se baseia nas propriedades dos textos (HARMAN, 1992), e levando em conta a hipótese de que algumas dessas propriedades podem se propagar para os criptogramas, é razoável o emprego de técnicas de classificação de textos como auxílio na descoberta de fraquezas nos algoritmos criptográficos atuais, embora esses algoritmos supostamente produzam uma distribuição quase uniforme dos símbolos dos criptogramas.

Em Carvalho (2006), encontramos um trabalho que já apresentou resultados relevantes na busca por fraquezas em algoritmos criptográficos com a utilização de técnicas de recuperação de informações. Nesse trabalho, foi realizado o agrupamento de criptogramas onde os grupos formados possuiam em comum o fato de terem criptogramas que foram cifrados com a mesma chave. Para o agrupamento, foi utilizado o método hieráquico aglomerativo de ligação simples (*single-link*). A medida do ângulo do co-seno foi utilizada para o cálculo da similaridade entre os criptogramas.

1.2 CARACTERIZAÇÃO DO PROBLEMA

A criptografia é tratada oficialmente pelo Brasil, assim como na Europa e nos Estados Unidos, como assunto de interesse para o desenvolvimento e segurança nacionais. Apesar disso, o Brasil não figura como um país desenvolvedor de produtos criptográficos (LAMBERT, 2004). Por ser assunto de grande relevância para o país, pode não ser seguro a utilização de produtos criptográficos desenvolvidos para o público em geral, sobretudo aqueles agregados a soluções de software, como as ferramentas de correio eletrônico e de *workflow*. Em Lambert (2004), pode ser visto um exemplo de quebra de sigilo com software desenvolvido pela empresa *Lotus Notes*.

Desta forma, o Plano Básico de Ciência e Tecnologia do Exército Brasileiro tem como um dos seus objetivos o desenvolvimento de soluções criptográficas, próprias do Exército

Brasileiro, visando à preservação do sigilo das informações transmitidas ou armazenadas em meios de tecnologia da informação (PBCT/EB, 2007).

Contudo, desenvolver produtos criptográficos requer meios para garantir a confiabilidade desses produtos. Por esse motivo alguns algoritmos criptográficos não são disponibilizados para o público em geral pelos seus desenvolvedores, os quais entendem que isso aumenta a sua segurança. Mas, têm-se relatos de algoritmos criptográficos apresentados ao público, os quais não resistiram à criptoanálise realizada na própria conferência onde foram apresentados (SCHNEIER, 1996), o que pode questionar a segurança de um sistema criptográfico baseada apenas no sigilo do seu algoritmo.

Assim, um conjunto de requisitos pode ser definido para avaliar a confiabilidade de um algoritmo criptográfico e, caso este algoritmo atenda a esses requisitos, ele pode ser considerado confiável. Contudo, não se tem conhecimento de nenhuma forma de avaliação que garanta a segurança desses algoritmos. Os testes propostos pelo NIST e pelo NESSIE estão relacionados à reprovação do algoritmo e a sua eficiência (CARVALHO, 2006) e (LAMBERT, 2004).

Um dos requisitos que pode ser utilizados para a garantia da confiabiliadade de um algoritmo criptográfico seria a inexistência de padrões nos criptogramas gerados por estes algoritmos. Como pode ser visto no trabalho de Carvalho (2006), foram detectados padrões em criptogramas gerados com os algoritmos DES e AES, com chaves de 64 e 128 *bits*, respectivamente, no modo de operação ECB. Estes padrões se caracterizaram pelo agrupamento dos criptogramas cifrados com a mesma chave.

Neste contexto, a presente dissertação expandiu aquele trabalho e explorou o uso de outras técnicas de agrupamento de textos, assim como das redes neurais artificiais para agrupamento e classificação de textos, com os seguintes objetivos:

a) Avaliar a influência de medidas de similaridade e distância no agrupamento de criptogramas;
b) Avaliar a correlação entre as medidas de similaridade e distância;
c) Avaliar o agrupamento de criptogramas utilizando outros métodos hierárquicos aglomerativos;
d) Avaliar o agrupamento de criptogramas, utilizando todas as medidas e métodos dos itens anteriores, para o algoritmo AES com chaves de 192 e 256 *bits*;

e) Avaliar o agrupamento de criptogramas, utilizando todas as medidas e métodos dos itens anteriores, para o algoritmo RSA com chaves de 64, 128, 256, 512 e 1024 *bits*; e

f) Avaliar a possibilidade de utilização de uma rede neural auto-organizável, baseada no mapa de Kohonen, aplicada ao agrupamento e classificação de criptogramas.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

No capítulo 2 é apresentada a criptografia e o seu contexto histórico, assim como, as cifras de blocos e os algoritmos criptográficos que serão alvos deste trabalho e os seus modos de operação.

O capítulo 3 faz um apanhado geral das técnicas de recuperação de informações utilizadas neste trabalho, considerando como as mesmas são aplicadas aos criptogramas.

No capítulo 4, introduzimos as redes neurais artificiais, especificamente o Mapa de Kohonen, aplicada ao agrupamento e classificação de criptogramas.

O capítulo 5 descreve as ferramentas desenvolvidas e utilizadas neste trabalho, detalhando as suas características.

No capítulo 6 são descritos os experimentos, feitas as avaliações e apresentados os resultados dos experimentos.

No capítulo 7 são feitas as conclusões e apresentadas as contribuições do presente trabalho e as possibilidades de trabalhos futuros.

2 CRIPTOGRAFIA CONTEMPORÂNEA

Criptografia, para Schneier (1996), é a arte e a ciência de manter mensagens seguras, criptoanálise é a arte e a ciência de quebrar[3] criptogramas e criptologia é o ramo da matemática que engloba ambas.

Para Menezes (1996), criptografia é o estudo de técnicas matemáticas relacionadas à segurança da informação.

As palavras criptologia e criptografia originam-se do grego e tem o mesmo significado: a arte e a ciência de criar métodos para garantir o sigilo de uma mensagem (BEUTELSPACHER, 1994). Contudo, como observarmos nos parágrafos anteriores, alguns autores fazem distinção entre as duas.

Aquele que deseja estabelecer um canal seguro de comunicações deve implementar um conjunto de medidas e contramedidas para impedir ou, pelo menos, para tornar mais difícil o trabalho de alguém que tenta obter um conhecimento protegido. Assim surgiu a criptografia, da necessidade de sigilo na transmissão de informações.

Um modelo para sistemas de comunicações pode ser com o apresentado na figura 2.1. Ele é composto por emissor, receptor, mensagem e canal.

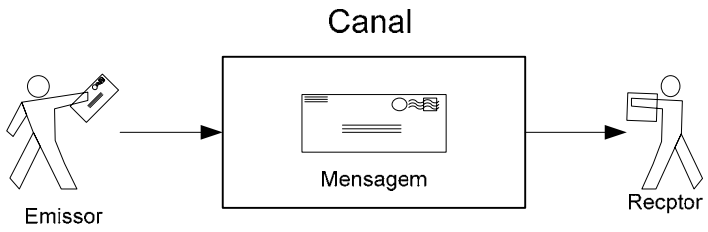


FIG. 2.1: Sistema de comunicações

Com a finalidade de proteger o conhecimento, devemos introduzir um item que represente um conhecimento adicional na transmissão da informação. Um conhecimento exclusivo do emissor e do receptor. Este conhecimento pode ser chamado de chave (BEUTELSPACHER, 1994). Deste modo, o nosso modelo de comunicações seguro, ficaria como apresentado na figura 2.2.

[3] Obter o respectivo texto em claro.

A chave apresentada na figura 2.2 é uma propriedade útil para classificar uma cifra. Desta forma, temos as cifras simétricas e assimétricas.

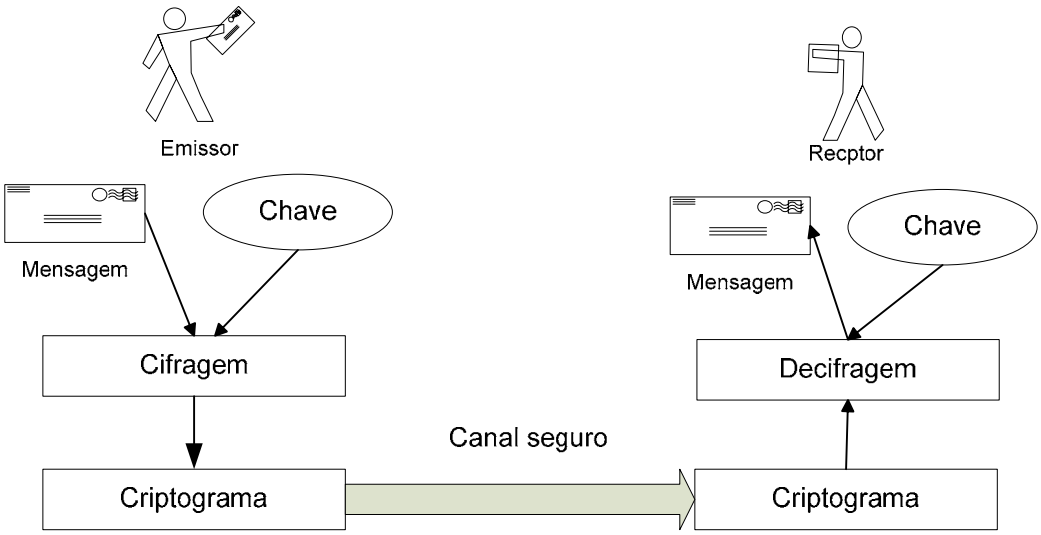


FIG. 2.2: Modelo de comunicações seguro

Nas cifras simétricas a chave utilizada para decifrar é igual à chave utilizada na cifragem ou é facilmente obtida a partir dela. Uma vez que a chave precisa ser mantida em segredo e deve ser transmitida por um canal seguro, estes sistemas são também conhecidos como sistemas de chave secreta. A segurança destes sistemas está no segredo da sua chave (SCHNEIER, 1996).

Nas cifras assimétricas a chave utilizada para decifragem é diferente da chave utilizada para cifragem ou não é facilmente obtida a partir dela. Estes sistemas são também chamados de sistemas de chave pública uma vez que a chave utilizada na cifragem pode ser pública (SCHNEIER, 1996), já que a segurança está relacionada ao segredo da chave de decifragem, isto é, a chave privada (LAMBERT, 2004).

Se por um lado os emissores das mensagens tentam prover o máximo sigilo às suas mensagens, por outro os adversários[4] buscam novas técnicas para conhecer o conteúdo dessas mensagens. Nesse contexto é que se desenvolve a criptoanálise, conjunto de técnicas usadas para obter o texto em claro sem o conhecimento da chave utilizada na função de cifrar

[4] Pessoas que não tem autorização para acesso ao conteúdo das mensagens.

(SINGH, 2000) e, alternativamente, pode ter o objetivo de descobrir a própria chave utilizada nesta função.

2.1 CONTEXTO HISTÓRICO

Durante toda a história, os grandes líderes, fossem militares, religiosos ou de nações, precisaram utilizar comunicação eficiente e segura, de maneira a formar alianças, divulgar as suas decisões e comandar os seus exércitos com a garantia de que as suas mensagens não seriam conhecidas por pessoas não autorizadas.

Essa ameaça de interceptação das mensagens, por pessoas não autorizadas, estimulou o desenvolvimento da criptografia (SINGH, 2001).

A escrita secreta se desenvolveu em dois ramos distintos: a esteganografia, que consiste em ocultar a mensagem, e a criptografia, que mistura o conteúdo da mensagem de maneira a tornar o seu significado incompreensível para as pessoas que não sabem como decifrá-la.

A criptografia utiliza-se de duas transformações: a transposição e a substituição. Na transposição é feita uma permutação dos caracteres da mensagem, de acordo com algum critério. A ordem original dos caracteres é modificada, de maneira que o texto fica embaralhado (KAHN, 1967). É importante notar que o conjunto de caracteres da mensagem não se altera, apenas as suas posições são modificadas (SINGH, 2001). Na substituição, cada caractere da mensagem é substituído por outro caractere, presente ou não na mensagem original.

Em resumo, na transposição os caracteres mantêm a sua identidade, mas perdem a sua posição. Na substituição, os caracteres mantêm a sua posição, mas perdem a sua identidade (KAHN, 1967) e (SINGH, 2000).

A criptografia por substituição, ou cifra de substituição, foi a que mais se desenvolveu e, assim, fornece um material mais interessante para exploração. Podemos considerar a operação desta cifra como a transformação de um caractere do alfabeto da língua original da mensagem em um caractere do alfabeto cifrado. Deste modo de operação, originou-se o nome cifra de substituição monoalfabética, ou simplesmente, cifra monoalfabética.

As cifras monoalfabéticas permaneceram seguras durante vários séculos, até que os árabes, baseados em técnicas matemáticas, estatísticas e lingüísticas inventaram a criptoanálise (KAHN, 1967). O estudo dos escritos sagrados do profeta Maomé, com a finalidade de identificar a cronologia das revelações feitas ao profeta e a legitimidade das

declarações atribuídas a ele, acabou revelando que algumas letras eram mais comuns do que outras (SINGH, 2000). Esta característica foi explorada pelo polímata árabe Al-Kindi, o qual batizou a técnica como criptoanálise por análise de freqüências. Esta técnica explorava a freqüência das letras do idioma original da mensagem comparada à freqüência dos símbolos dos caracteres da mensagem cifrada. Para o uso efetivo dessa técnica, era necessário conhecer o idioma de origem da mensagem cifrada, assim como, possuir uma tabela de freqüência das letras no alfabeto de origem.

A descoberta da criptoanálise trouxe insegurança para aqueles que necessitavam enviar mensagens cifradas. Estes conviviam com o medo de que suas mensagens fossem interceptadas e decifradas pelos adversários.

Essa nova demanda por segurança levou a criação da cifra polialfabética, a qual foi descrita pela primeira vez por Leon Alberti, um polímata florentino do século XV (BEUTELSPACHER, 1994) e (SINGH, 2000). Na cifra de Alberti, um caractere do alfabeto da língua original da mensagem poderia ser transformado em um caractere de um ou mais alfabetos cifrados, de maneira alternada, isto é, a cada transformação de caracteres, um dos alfabetos era escolhido, segundo algum critério, para fornecer o caractere cifrado.

Pouco mais de um século depois, Blaise de Vigenère, desenvolveu uma cifra polialfabética, a qual ficou conhecida como cifra de Vigenère. Esta c.2(o) TcC(o m)8.9ea m -11(to)

(BEUTELSPACHER, 1 994s.

cifra de Vigenere era cos

se constituiu em um marco na história da criptografia. A Enigma foi amplamente utilizada pelas forças alemães durante a Segunda Guerra Mundial.

Embora fosse uma máquina formidável (SINGH, 2000) e mesmo sofrendo diversas modificações para aumentar a segurança da cifra que a mesma implementava, a sua cifra foi quebrada pelo polonês Mirian Rejewski, o qual usou além de seus conhecimentos matemáticos, um conjunto de réplicas da máquina Enigma associadas a outros mecanismos, denominadas Bombas.

Com o início da Segunda Guerra Mundial, os alemães aperfeiçoaram a Enigma a tal ponto que não foi mais possível para Rejewski realizar a decifragem das mensagens alemães. Assim, os poloneses enviaram aos ingleses todo o trabalho feito até então sobre a Enigma. Este trabalho surpreendeu os criptoanalistas ingleses pelos resultados obtidos e os motivou na sua guerra particular contra a Enigma. De fato, esta guerra foi vencida pelos ingleses, graças ao matemático Alan M. Turing, o qual projetou uma máquina, também denominada Bomba, a qual podia decifrar as mensagens alemãs. Ainda relevante foi a contribuição de Max Newman, o qual desenvolve o Colossus (CARVALHO, 2006), considerada como um dos primeiros computadores modernos.

No front do pacífico, os japoneses possuíam a sua própria máquina de cifras, conhecida como Púrpura. Os norte-americanos conseguiram decifrar a Púrpura, mudando o curso da guerra naquele front. Digno de registro também é o Código Navajo usado na campanha do Pacífico, o qual era baseado no idioma da tribo indígena de mesmo nome. Este código permaneceu indecifrável (SINGH, 2000).

Muitos acontecimentos históricos, principalmente as guerras, mostraram a importância da criptografia. Maìs recente ainda, as instituições comerciais entenderam também o valor da criptografia para o sigilo de suas operações. Assim, soluções criptográficas polularam ao redor do mundo, até que na década de 1970 o governo norte-americano adotou um algoritmo desenvolvido pela IBM, o qual viria a chamar-se Data Encryption Standard (DES).

Após mais de 20 anos de uso do DES, o governo norte-americano entendeu que havia a necessidade do estabelecimento de um novo padrão criptográfico. Para tanto, lançou uma competição, a qual foi vencida pelo o algoritmo Rijndael, o qual, a partir de 2001, passou a ser conhecido como Advanced Encryption Standard (AES).

Outros algoritmos como o RSA (Rivest, Shamir e Adleman), o qual recebeu este nome em homenagem aos seus criadores, são igualmente importantes para a criptografia. Estes três algoritmos serão detalhados nas próximas sessões deste capítulo.

Através dos parágrafos anteriores, verificamos que há uma guerra constante entre criadores de códigos e os seus pretensos decifradores. Atualmente, os criadores de códigos estão ganhando a batalha, uma vez que não existe nem um ataque prático e significativo às cifras padrões existentes. Ressalvamos que o assunto criptologia é tratado como assunto de

necessitam que todos os blocos possuam o mesmo tamanho, isto é, 64 bits. Assim, é necessário utilizar um algoritmo de ***padding***[5], para completar esse último bloco.

2.2.1.1 MODO *ELETRONIC CODEBOOK* (ECB)

Neste modo de operação o mesmo bloco de mensagem[6] é sempre cifrado para o mesmo bloco de criptograma, o que pode permitir a criação de um livro de códigos. Disto vem o nome *codebook*. Embora, isso possa representar uma fraqueza, se considerarmos um bloco 64 *bits*, o de livro de códigos terá 2^{64} entradas. Além disso, para cada chave utilizada teremos um livro de código diferente.

Os processos de cifragem e decifragem são mostrados nas figuras 2.3 e 2.4, respectivamente.

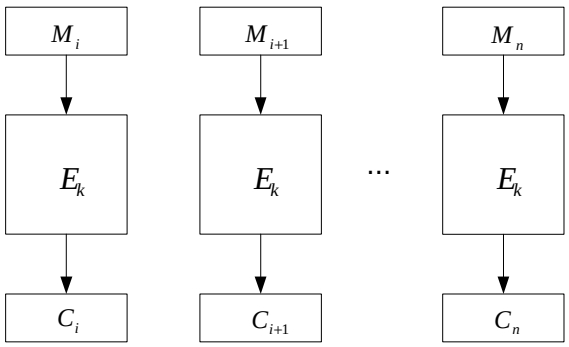


FIG. 2.3: Processo de cifragem no modo ECB

O modo ECB pode ser visto como uma cifra monoalfabética, já que um bloco de texto em claro pode ser traduzido para apenas um bloco de texto cifrado.

[5] "Encher com itens falsos", i.e., completar o bloco com algum padrão pré-estabelecido. por exemplo, com zeros, com uns ou com ambos.
[6] Neste trabalho usaremos os termos mensagens e texto em claro como sinônimos.

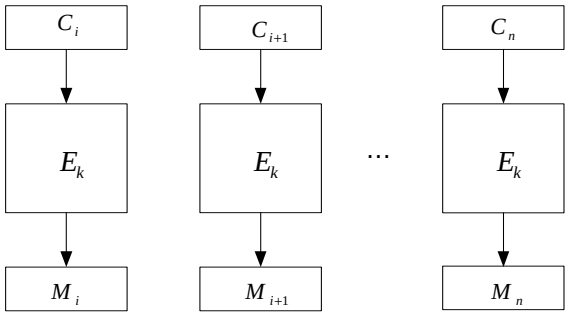


FIG. 2.4: Processo de decifragem no modo ECB

2.2.1.2 MODO *CIPHER BLOCK CHAINING* (CBC)

Este modo de operação adiciona um mecanismo de retroalimentação à cifra, isto é, o bloco cifrado anteriormente é usado como entrada no processo de cifragem do bloco atual. Neste processo, antes de o texto em claro ser cifrado, é realizada uma operação *XOR* entre este texto e o bloco cifrado anteriormente. Assim, é criada uma cadeia de dependência entre o bloco atual e todos os blocos anteriores a ele.

A cifragem é dada pela equação $C_i = E_k(M_i \oplus C_{i-1})$ (FIG. 2.5) e o decifragem é pela equação $M_i = C_{i-1} \oplus D_k(C_i)$ (FIG. 2.6), Onde C_i é o $i-ésimo$ bloco do criptograma, E_k é a função de cifrar, D_k é a função de decifrar, k é a chave utiliza para cifrar e decifrar, M é o $i-ésimo$ bloco da mensagem e C_{i-1} é o bloco cifrado resultante do processo de cifragem imediatamente anterior.

Assim como no modo ECB, neste modo um texto em claro, cifrado com a mesma chave vai produzir sempre o mesmo criptograma (SCHNEIER, 1996) e (MENEZES, 1996). A diferença em relação ao modo ECB é que pode ser usado um vetor de inicialização, o qual troca o primeiro bloco da cifrado por um valor variável. Desta forma, o modo CBC equivale a uma cifra polialfabética, onde um bloco de texto em claro pode ser traduzido para um de muitos blocos de texto cifrado, dependendo do valor do vetor de inicialização.

Nas figuras 2.3 e 2.4 são apresentados, respectivamente, os processos de cifragem e decifragem, com o uso do vetor de inicialização.

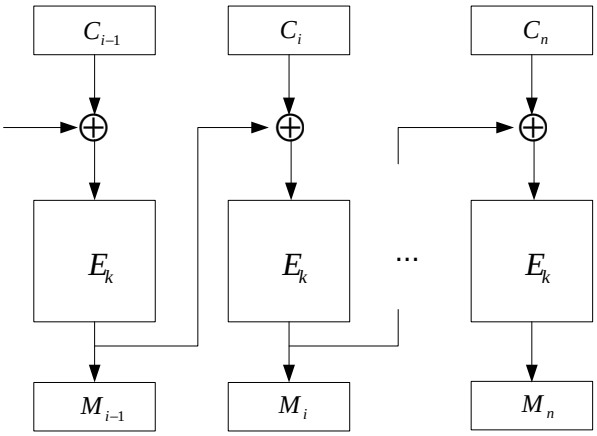


FIG. 2.5: Processo de cifragem no modo CBC

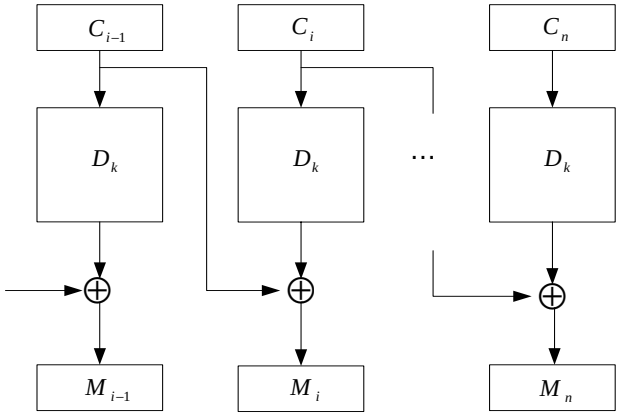


FIG. 2.6: Processo de decifragem no modo CBC

2.2.2 O ALGORITMO DATA ENCRYPTION STANDARD (DES)

Em meados da década de 1970, o governo Norte-Americano, por meio do NBS, atual NIST, adotou o DES como padrão de criptografia para informações que, embora sensíveis, não requeriam alto grau de sigilo (MENEZES, 1996) e (LAMBERT, 2004).

Esta adoção se deveu ao fato de que as pesquisas e os meios criptográficos para fins não-militares encontravam-se sem uma direção a ser seguida, embora o potencial de uso da criptografia em aplicações comerciais ou industriais fosse muito bom, dado ao desenvolvimento do poder computacional.

2.2.2.1 CARACTERÍSTICA DO DES

O DES é uma cifra de blocos de 64 *bits*, baseada na estrutura Feistel, cujo processo de cifragem é realizado em 16 iterações ou rodadas. O tamanho da chave para a cifragem é de 64 *bits*, mas a chave real é de 56 *bits*. Os oito bits restantes não são usados, a princípio, pelo algoritmo, mas podem ser usados para detecção de erro (NIST, 1999).

2.2.2.2 A ESTRUTURA FEISTEL

Na estrutura de Feistel, um bloco de comprimento n é divido em duas partes de comprimento $n/2$ cada uma. O tamanho n deve ser par. Seja i a $i-ésima$ iteração. Sejam B_i o bloco que será cifrado com a chave k_i, f a função que será utilizada na cifragem e L_i a parte esquerda do bloco B_i e R_i a parte direita do bloco B_i. Assim, a cifragem é dada por (figura 2.7):

$$L_i = R_{i-1}$$

$$R_i = L_{i-1} \oplus f(R_{i-1}, K_i)$$

Onde $1 \leq i \leq nr$, e nr número total de iterações.

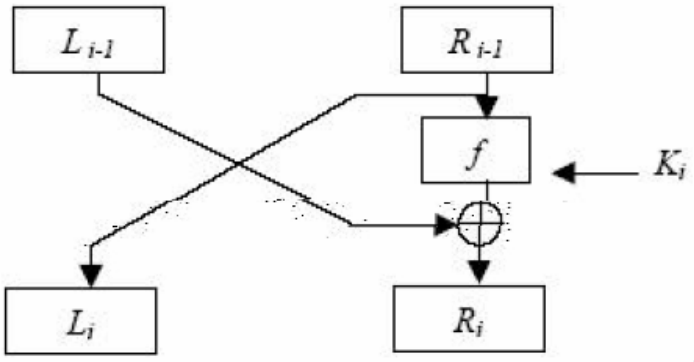


FIG. 2.7: Descrição da operação de cifragem na estrutura Feistel

A decifragem se baseia na propriedade de simetria da operação lógica *XOR* (ou exclusivo). Esta operação equivale à soma módulo dois. Sejam B o bloco, L a parte

esquerda do bloco B e R a parte direita do bloco B, onde cada parte possuem o mesmo tamanho. Então, $(L \oplus R) \oplus R = L$. Logo a operação de decifragem é definida por (figura 2.8):

$$R_{i-1} = L_i$$

$$L_{i-1} = R_i \oplus f(R_{i-1}, K_i)$$

Onde $1 \leq i \leq nr$, e nr número total de iterações.

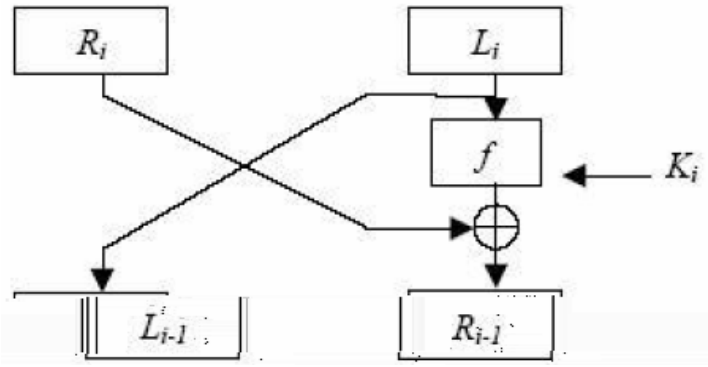


FIG. 2.8: Descrição da operação de decifragem na estrutura Feistel

2.2.2.3 FUNCIONAMENTO DO DES

O esquema de funcionamento do DES começa com uma permutação inicial do bloco de entrada. Após isto, o resultado é dividido em duas partes iguais e são realizadas as 16 rodadas, conforme a figura 2.9. Ao final das rodadas, as partes são concatenadas e submetidas a uma permutação final, que é a operação inversa da permutação inicial, criando um bloco cifrado.

A força desta cifra reside na função f utilizada. Esta função é utilizada durante as iterações (figura 2.7 e figura 2.9) e pode ser definida por $f(R_{i-1}, K_i) = P(S(E(R_{i-1}) \oplus K_i))$. Inicialmente é realizada uma operação de EXPANSÃO na metade R, a qual expande a entrada de 32 bits para 48 bits. É realizada uma operação XOR com a subchave da rodada. O resultado é submetido à operação de SUBSTITUIÇÃO a qual utiliza oito caixas-S[7] para transformar a entrada de 48 bits em uma saída de 32 bits. Por fim, é realizada uma operação

[7] S-box

de PERMUTAÇÃO sobre o bloco de 32 bits recebidos. Para maiores detalhes sobre essas operações, consultar NIST (1999).

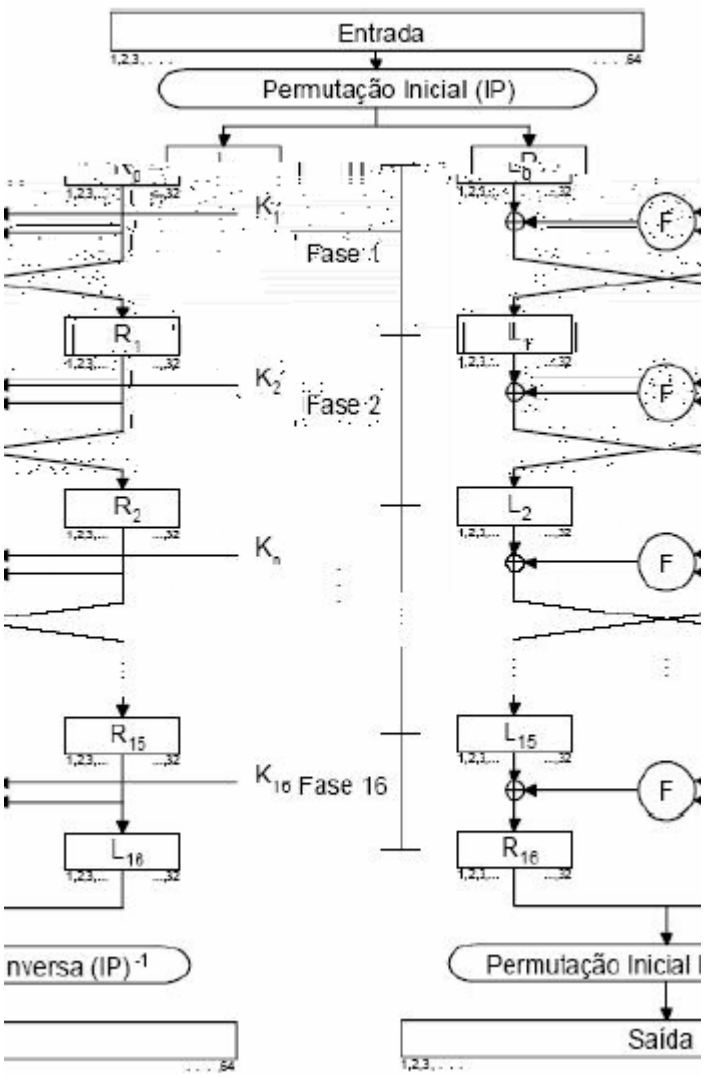


FIG. 2.9: Descrição do funcionamento do DES (CARVALHO, 2006)

2.2.3 O ALGORITMO *ADVANCED ENCRYPTION STANDARD* (AES)

O avanço da tecnologia dos computadores permitiu que fossem exploradas fraquezas no DES, tornando necessário o estabelecimento de um novo padrão criptográfico.

Assim, em janeiro de 1997 (LAMBERT, 2004) o NIST lançou um concurso para escolha de um novo padrão de cifra. Ao final da primeira etapa do concurso, cinco algoritmos foram

escolhidos: MARS (IBM), RC6 (RSA), Rijndael (John Daemen e Vincent Rijmen), Serpent (Ross Anderson, Eli Biham e Lars Knudsen) e TwoFish (Bruce Schneier, John kelsey e outros), sendo o vencedor do concurso o Rijndael, tornando-se conhecido a partir de então como AES (NIST, 2001).

2.2.3.1 CARACTERÍSTICAS DO AES

O AES é uma cifra simétrica de blocos, a qual trabalha com blocos de dados de 128, 192 e 256 *bits*, com chaves que podem ter os mesmos tamanhos dos blocos (LAMBERT, 2004), embora no padrão estabelecido pelo NIST (2001) seja adotado somente o bloco de dados com 128 *bits*.

O processo de criptografia pode ocorrer em 10, 12 ou 14 iterações ou rodadas, para as chaves de 128, 192 e 256 *bits*, respectivamente.

2.2.3.2 FUNCIONAMENTO DO AES

O processo de cifragem do AES é realizado por meio de seis operações ou transformações: *AddRoundKey*, *SubBytes*, *ShiftRows*, *MixColumns*, *KeyExpansion* e *RotWord*. Neste trabalho, utilizaremos a terminologia sugerida por Lambert (2004), conforme a tabela 2.1. Diferente do DES, o AES não utiliza a Estrutura Feistel. Assim, no processo de decifragem, são utilizadas operações inversas das operações listadas acima (NIST, 2001).

TAB. 2.1 – Terminologia alternativa para as operações do AES

Sugerida	**FIPS 197**
SOMASUBCHAVE	*AddRoundKey*
SUBSTITUIÇÃO	*SubBytes*
PERMUTAÇÃO	*ShiftRows*
MULTIPLICAÇÃO	*MixColumns*
EXPANSÃO	*KeyExpansion*
ROTAÇÃO	*RotWord*

A transformação SOMASUBCHAVE é uma soma módulo dois, i.e., uma operação lógica “OU EXCLUSIVO” realizada *bit* a *bit*, entre a chave e o bloco da mensagem.

A SUBSTITUIÇÃO é uma operação realizada *byte* a *byte*, na qual utilizamos uma matriz chamada de caixa de substituição ou caixa-S, onde é feito o cruzamento entre a matriz e o *byte* do bloco. A metade esquerda do *byte* indica a linha da matriz e a metade direita indica a coluna da matriz.

Na PERMUTAÇÃO simplesmente realizamos uma troca na posição dos *byte* de um bloco. Seja B, um bloco de 128 *bits* e P a operação de permutação. Então, após a operação, teríamos o seguinte arranjo:

$$P(B_{128}) = (b_0, b_5, b_{10}, b_{15}, b_4, b_9, b_{14}, b_3, b_8, b_{13}, b_2, b_7, b_{12}, b_1, b_6, b_{11})$$

A operação de MULTIPLICAÇÃO é realizada por meio da multiplicação de uma matriz constante de 32 x 32 *bits*. Assim, um bloco de 128 *bits* de entrada é dividido em quatro palavras de 32 *bits*. Cada palavra é então multiplicada pela matriz. As palavras resultantes da multiplicação são concatenadas para forma a saída da operação.

A operação de MULTIPLICAÇÃO pode ser encontrada de maneira detalhada em Lambert (2003), onde também encontramos uma proposta de simplificação para a mesma.

EXPANSÃO consiste em gerar NR+1 subchaves que serão utilizadas pelas operações de SOMASUBCHAVE, ou seja, uma subchave para cada vez que a operação SOMASUBCHAVE for executada. A primeira subchave é a própria chave fornecida na entrada.

A ROTAÇÃO faz parte da EXPANSÃO é consiste na realização de um deslocamento cíclico de *bytes* à esquerda. Assim, seja B, um bloco de 128 bits, o qual foi dividido em quatro palavras de 32 bits. Seja R a operação de ROTAÇÃO. Então, temos:

$$B_{128} = (p_0, p_1, p_2, p_3),\ R(p_0, p_1, p_2 p_3) = (p_1, p_2, p_3, p_0)$$

O fluxo de execução do AES é descrito na figura 2.10. O algoritmo recebe uma chave a qual é submetida a uma operação de EXPANSÃO. A finalidade desta operação é gerar um total de NR+1 subchaves que serão utilizadas ao longo do processo de cifragem, onde NR é o número de rodadas. Em seguida, é realizada a operação SOMASUBCHAVE, a qual realiza

uma soma módulo dois entre o resultado da EXPANSÃO e a mensagem original. Nesta primeira etapa, a chave é a própria chave original recebida na entrada.

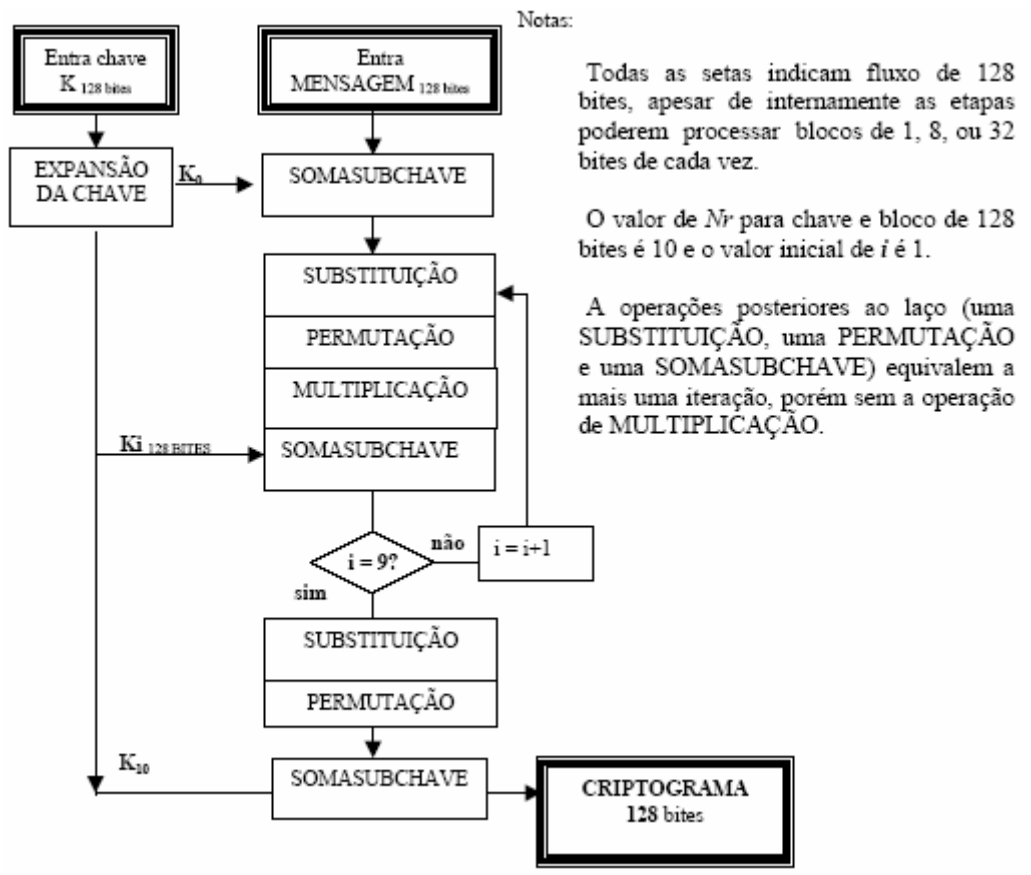


FIG. 2.10 – Fluxo de execução do AES (LAMBERT, 2004)

Na próxima etapa, iniciamos o laço, onde serão realizadas *NR-1* operações de SUBSTITUIÇÃO, PERMUTAÇÃO, MULTIPLICAÇÃO e SOMASUBCHAVE.

Após *NR-1* rodadas, o algoritmo realiza uma operação de SUBSTITUIÇÃO, PERMUTAÇÃO e SOMASUBCHAVE, sendo finalmente gerado o criptograma.

2.2.4 O ALGORITMO *RIVEST-SHAMIR-ADLEMAN* (RSA)

O RSA[8] é um algoritmo de criptografia assimétrica, que pode ser usado tanto em sistemas criptográficos quanto para assinaturas digitais (SCHNEIER, 1996) e é baseado na

[8] O RSA possui este nome em homenagem aos seus criadores Ron Rivest, Adi Shamir e Leonard Adleman.

dificuldade computacional de se determinar os fatores primos de um número inteiros muito grandes[9] (TERADA, 2000).

2.2.4.1 CARACTERÍSTICAS DO RSA

A criptografia assimétrica é diferente da criptografia simétrica, pois utiliza um par de chaves: uma privada, que deve ser mantida em segredo e outra pública, que pode ser disponibilizada para o público em geral. É conhecida também como criptografia de chave pública, sendo criada por Whitfield Diffie e Martin Hellman, em 1975. Este sistema trabalha com problemas matemáticos difíceis[10] de resolver. A segurança do sistema está baseada na resistência de uma chave em ser descoberta, dada a outra chave (SCHNEIER, 1996) e na dificuldade de deduzir-se o texto original, dado o texto cifrado.

O RSA é em geral mais lento do que os sistemas de criptografia simétrica, sendo tipicamente utilizado em conjunção com estes algoritmos em sistemas de comunicações para troca eletrônica de chave.

2.2.4.2 FUNCIONAMENTO DO RSA

O RSA trabalha sobre o princípio da dificuldade de se fatorar números inteiros muito grandes. Assim, é necessário encontrar dois números primos *p* e *q* grandes e de igual tamanho, para que se obtenha a máxima segurança (SCHNEIER, 1996). Depois, calcula-se o seu produto, $n = pq$. O próximo passo é encontrar um valor *e*, aleatório em $Z^*_{\varphi(n)}$, ou seja, um inteiro entre 1 e $\varphi(n)$, que seja relativamente primo com $\varphi(n)$, onde φ é a função de Euler[11] e $\varphi(n)$ é o número de inteiros relativamente primos com *n* (SCHEINERMAN, 2003). Note que $\varphi(n) = (p-1)(q-1)$ (TERADA, 2000).

O passo seguinte é o cálculo de um valor *d*, o qual é relativamente primo com *n* (SCHNEIER, 1996), tal que $d = e^{-1}$ em $Z^*_{\varphi(n)}$. Ou seja, *d* é o inverso de *e* no contexto dos números relativamente primos com $(p-1)(q-1)$.

[9] Para os dias atuais, número entre 100 e 200 dígitos ou mais.
[10] Tarefa que não pode ser realizada em tempo oportuno, com a tecnologia atual dos computadores (LAMBERT, 2004).
[11] A função de Euler e definida pelo seguinte teorema: Sejam *n* um inteiro positivo e *a* um inteiro relativamente primo com *n*. Então $a^{\varphi(n)} \equiv 1 (\mathrm{mod}\, n)$ (SCHEINERMAN, 2003).

Note que $d = e^{-1} \bmod((p-1)(q-1))$ é igual a $d = e^{-1}$ em $Z^*_{\varphi(n)}$, já que $\varphi(n) = (p-1)(q-1)$.

Se na escolha aleatória, for selecionado o $e = n$ em $Z^*_{\varphi(n)}$ a chave privada será igual a 1.

Dadas as definições abaixo transcritas de Scheinerman (2003):

"Sejam *n* um inteiro positivo e $a \in Z_n$. O inverso de *a* é um elemento $b \in Z_n$ tal que $a \otimes b = 1$. Um elemento de Z_n que tenha um inverso é chamado de invertível."

"Sejam *a* e *b* inteiros. Dizemos que *a* e *b* são relativamente primos se e somente se $mdc(a,b) = 1$."

"Sejam *a* e *b* inteiros. Existem inteiros *x* e *y* tais que $ax + by = 1$ se e somente se *a* e *b* são relativamente primos."

"O cálculo do inverso é realizado determinando-se o valor de *x* e *y*, por meio do algoritmo de Euclides estendido[12], onde $a = ((p-1)(q-1))$ e $b = e$."

Podemos definir a chave pública como *e* e *n*, a chave privada como *d* a função de cifrar pela fórmula 2.1 e a função de decifrar pela fórmula 2.2.

$$c = m^e \bmod n \qquad \text{Fórmula 2.1}$$

$$m = c^d \bmod n \qquad \text{Fórmula 2.2}$$

Onde *m* é a mensagem a ser cifrada e *c* é o criptograma resultante.

Após a criação das chaves, os números *p* e *q* devem ser descartados sem serem revelados (SCHNEIER, 1996), (TERADA, 2000) e (SCHEINERMAN, 2003).

2.2.4.3 UM EXEMPLO DE CIFRAGEM E DECIFRAGEM COM O RSA

O processo de cifrar de uma mensagem por meio do algoritmo RSA pode ser efetuado por meio dos passos abaixo:

1. Sejam $p = 71$ e $q = 37$. Então, $n = pq = 2627$.

2. Calcular $\varphi(n) = (p-1)(q-1)$, 70 x 36 = 2520.

---

[12] Um exemplo de execução do algoritmo de Euclides estendido pode ser visto em Souza

3. Escolher *e* aleatoriamente de Z^*_{2520}. Seja $e = 23$.

4. Calcular $d = e^{-1}$ em Z^*_{2520}. Então, $d = 767$.

5. Aplicar a fórmula 2.1 sobre a mensagem *m*. Seja $m = 153$. Então, $c = 153^{23} \bmod 2627 = 1093$.

Para decifrar, aplicamos a fórmula 2.2, ou seja, $m = 1093^{767} \bmod 2627 = 153$.

É importante notar que a mensagem *m* não pode ser maior do que *n*, senão o processo de cifragem e decifragem falhará. Por exemplo, utilizando as chaves que já foram criadas, cifrar a mensagem $m = 4556$. Assim, $m > n$.

Aplicando a fórmula 2.1, $c = 4556^{23} \bmod 2627 = 649$.

Para decifrar, aplica-se a fórmula 2.2, $m = 649^{767} \bmod 2627 = 1929$.

Percebe-se que a mensagem não foi decifrada.

2.2.4.4 ALGUMAS QUESTÕES RELEVANTES SOBRE O RSA

A escolha dos números primos *p* e *q*, tem importância crucial na segurança do RSA. Uma vez que a segurança se baseia na dificuldade de fatorar *n*, sendo $n = pq$, devemos ter *p* e *q* bem grandes. Além disso, é necessário que o número de dígitos do resto da subtração $p - q$, seja grande também, do contrário será fácil fatorar *n* usando o algoritmo de Fermat. Para maiores detalhes, consulte (COUTINHO, 2000).

Outra questão relevante é a escolha do número *e*, já que este número deve ser relativamente primo com $\varphi(n)$, devemos ter cuidado na escolha do mesmo para garantir essa propriedade e também para não criarmos um número *d* que seja fácil de ser descoberto. Schneier (1996) indica que as escolhas mais comuns para este número são 3, 17 e 65537.

3 RECUPERAÇÃO DE INFORMAÇÕES APLICADA AOS CRITPOGRAMAS

No mundo contemporâneo, as informações são produzidas com uma velocidade muito superior à capacidade de percepção e absorção do homem. Com o surgimento da internet e com a consolidação do seu uso, pelo menos dois outros fatores de complexidade foram adicionados ao tratamento da informação: a divulgação em larga escala e a distribuição ao longo da grande rede mundial.

Além da dimensão do tratamento pessoal da informação, cada vez mais os ambientes organizacionais, sejam agências governamentais, indústrias ou empresas prestadoras de serviços, são confrontados com o problema de tratar uma imensa quantidade de informações essenciais para o seu trabalho, especialmente as informações textuais (MANNING, 2003).

Uma maneira de abordamos estes problemas é por meio do uso dos Sistemas de Recuperação de Informações (sistemas RI). Os sistemas RI comparam uma declaração formal de necessidade de informação, chamada consulta, com as informações disponíveis em uma base de dados (FRAKES, 1992). No contexto da Recuperação de Informações (RI) esta base de dados é composta por um conjunto de objetos, que representam documentos. Os documentos são arquivos digitais, os quais são geralmente compostos de informações textuais, mas podem conter também, imagens, gráficos, fórmulas matemáticas e outras formas de representar informações (FRAKES,de da5(s s de)]TJ-68.9411 -1.7224T*-0.0002 Tc-0.0002 T

características estatísticas da linguagem, especificamente tratando da freqüência dos caracteres. Logo, esta abordagem nos parece adequada para o tratamento dos criptogramas.

3.1 PROPRIEDADES DOS SISTEMAS RI[13]

Os sistemas RI podem ser classificados por meio de um conjunto de propriedades, as quais caracterizam estes sistemas. As seguintes propriedades podem ser utilizadas (FRAKES, 1992):

- *Modelo conceitual*: esta propriedade diz respeito ao tipo de modelo que será utilizado para modelar a coleção de objetos. Dentre os mais conhecidos temos o modelo booleano (YATES, 1999), o modelo de espaço de vetores e o modelo probabilístico (HARMAN, 1992) e (YATES, 1999);
- *Estrutura do arquivo*: é a estrutura na qual ficará armazenada a coleção de objetos. É importante notar que a utilização dessa estrutura pressupõe um processamento prévio da coleção de entrada, já que esta coleção é composta por um conjunto de arquivos digitais com uma estrutura própria, as quais devem ser apropriadamente tratadas e as informações relevantes armazenadas na estrutura escolhida. Exemplos de estrutura são os arquivos invertidos (HARMAN, 1992a) e arquivos de assinaturas (FALOUTSOS, 1992);
- *Operações de consulta*: uma consulta é a formalização de um pedido de informação a um sistema RI. As operações de consulta são as maneiras pelas quais podemos formalizar uma consulta. Como exemplo, temos uma consulta booleana, a qual utiliza operadores lógicos (WARTIK, 1992) para formalizar a consulta;
- *Operações sobre os termos*[14]: estas operações são realizadas sobre os termos com a finalidade de melhorar as respostas a uma consulta. Dentre estas operações, podemos destacar a lematização (FRAKES, 1992b), a remoção de *stoplist* (FOX, 1992) e a pesagem (HARMAN, 1992); e

[13] A partir deste ponto, se nada for dito em contrário, usaremos sistemas RI para definir os sistemas utilizados para recuperação de informações textuais.

[14] Termo é a maneira genérica de nos referirmos a unidade básica de para a recuperação de informações. Quando estivermos falando de textos os termos são palavras. Quando estivermos falando sobre criptogramas, os termos representam um bloco de um criptograma.

- *Operações sobre os documentos*: são operações realizadas sobre os documentos, normalmente para ordenar (HARMAN, 1992), classificar ou agrupar (JAIN, 1999) e (RASMUSSEN, 1992) os documentos recuperados em uma consulta;

Das informações textuais, podemos depreender o assunto por meio das palavras que compõem um texto. Outro dado importante é que quanto maior a interseção entre dois documentos, isto é, quanto mais palavras em comum eles possuírem, maior a chance de eles

serão considerados documentos, é necessário definir uma unidade básica para tratamento dos criptogramas, considerando apenas o aspecto léxico.

As cifras de bloco geram blocos cifrados compostos por zeros e uns, os quais, quando concatenados, formam um criptograma. Desta forma, pode-se dizer que o alfabeto de um criptograma é um alfabeto binário. Como estão sendo consideradas as cifras de bloco, a unidade básica de recuperação de criptogramas[15] pode ser um bloco, onde o tamanho deste bloco dependerá do algoritmo utilizado. Por exemplo, pode-se ter um bloco de 64 bits para o DES e um bloco de 128 bits no caso do AES (CARVALHO, 2006).

O modelo proposto busca identificar a quantidade de palavras sem repetição da coleção de documentos e a distribuição dessas palavras pelos documentos, para depois compará-los e verificar o grau de similaridade entre estes documentos, isto é, o quanto os documentos são parecidos. Assim, quanto mais similares dois documentos são, maior a chance de pertencerem ao mesmo grupo. A justificativa para a escolha de um bloco como unidade básica para a recuperação dos criptogramas é a mesma: quanto mais blocos em comum dois criptogramas possuírem, maior a chance de pertencerem ao mesmo grupo.

Existe ainda a tarefa de definir uma característica principal para identificar os grupos de criptogramas formados.

Quando os documentos são considerados, pode-se ter como característica principal o assunto ao qual pertence o documento. Este assunto pode estar escrito em uma língua qualquer e utilizando um alfabeto adequado a esta língua. Como foi visto anteriormente, em cifras de blocos, no modo ECB, se um mesmo bloco de texto em claro é cifrado duas vezes com a mesma chave, os dois blocos de criptogramas resultantes deste processo serão iguais. Considerando que uma chave qualquer submetida com um texto em claro a um algoritmo criptográfico determina uma linguagem particular, este processo dará origem a criptogramas escritos nesta linguagem (CARVALHO, 2006). Assim criptogramas produzidos com uma mesma chave serão mais similares, uma vez que compartilharão o alfabeto (binário) e a mesma linguagem (determinada pela chave), logo devem possuir o mesmo conjunto de elementos léxicos.

Outro fato importante é que a diversidade de termos gerados para as cifras de bloco é muito grande, por exemplo: 2^{64} para blocos de 64 bits e 2^{128} para blocos de 128 bits. Isto

[15] A partir deste ponto, vamos usar recuperação de criptogramas como o nome genérico de um sistema RI aplicado a criptogramas.

significa que dificilmente existirão blocos de criptogramas iguais em criptogramas diferentes que tenham sido gerados por chaves distintas (CARVALHO, 2006).

3.1.1 MODELO DE ESPAÇO DE VETORES

Neste modelo, a coleção de documentos é tratada como um espaço de vetores, onde cada documento representa um vetor de *n-dimensões*, sendo *n* o número de palavras da coleção de documentos, sem repetição (HARMAN, 1992). Desta forma, cada palavra representa um eixo no espaço vetorial e, consequentemente, um ponto no espaço determinará um documento (figura 3.1).

É muito provável que os documentos da coleção não possuam todos os eixos. Desta forma, a posição do vetor relativa à palavra recebe o valor um, se o documento que o vetor representa possuir a palavra, ou zero, caso contrário.

Na figura 3.1 pode-se notar que os documentos 1, 2 e 3 são pontos no espaço, representados pelas coordenadas (X, Y, Z), (X, 0, Z) e (0, Y, Z), respectivamente.

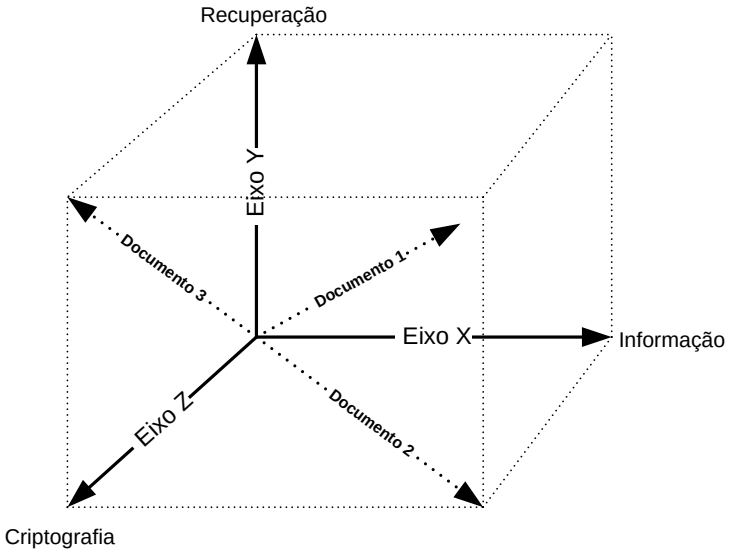


FIG. 3.1: Modelo de espaço de vetores

A quantidade de palavras sem repetição é igual a três (Informação, Recuperação, Criptografia), ou seja, $n = 3$. Temos então um espaço vetorial de três dimensões e os seguintes vetores:

Documento 1 (Informação, Recuperação, Criptografia) = vetor 1(1, 1, 1).

Documento 2 (Informação, 0 , Criptografia) = vetor 2(1, 0, 1).

Documento 3 (0 , Recuperação, Criptografia) = vetor 3(0, 1, 1).

Uma descrição formal seria: sejam d_1 e d_2, dois documentos. O valor relacionado à $d_{i,1}$, onde *i* representa a *i-ésima* palavra do documento 1, será 1, se o documento 1 contém a palavra, ou 0 caso contrário. O valor relacionado à $d_{i,2}$, onde *i* representa a *i-ésima* palavra do documento 2, será 1, se o documento 2 contém a palavra, ou 0 caso contrário. Então, o vetor para o documento 1 é definido como $\vec{d}_1 = (d_{1,1}, d_{2,1}, ..., d_{n,1})$, assim como o vetor para o documento 2 é representado por $\vec{d}_2 = (d_{1,2}, d_{2,2}, ..., d_{n,2})$, onde n é o número de palavras únicas na coleção de documentos. A coleção de documentos é denotada por $D = (d_1, d_2, ..., d_n)$.

3.1.1.1 MEDIDAS DE SIMILARIDADE E DISTÂNCIA

Para realizar o agrupamento de objetos de uma coleção de entrada, é necessário determinar o grau de associação entre esses objetos. Este grau de associação pode ser uma medida de similaridade ou dissimilaridade (RASMUSSEN, 1992). A dissimilaridade pode ser também chamada de distância (TEKNOMO, 2006).

A similaridade é o valor que indica a força da associação entre dois objetos. Estes valores estão normalmente entre -1 e 1 (TEKNOMO, 2006). Para os objetivos deste trabalho, é necessário normalizar o limite destes valores para 0 e 1, uma vez que não existe a possibilidade de similaridades negativa entre os objetos tratados, sejam documentos, sejam criptogramas. A dissimilaridade mede a distância entre dois objetos, isto é, o quanto dois objetos são divergentes (TEKNOMO, 2006).

A relação entre a similaridade e a dissimilaridade, considerando o limite normalizado, pode ser dada como segue: sejam *i* e *j* dois objetos de uma coleção de entrada e seja $d_{i,j}$ o valor da dissimilaridade entre esses dois objetos. Então, o valor da similaridade $s_{i,j}$ entre esses dois objetos é dado pela fórmula 3.1.

$$s_{i,j} = 1 - d_{i,j} \qquad \text{Fórmula 3.1}$$

A máxima similaridade ocorre quando $s_{i,j} = 1$ e a mínima dissimilaridade ou a menor distância ocorre quando $d_{i,j} = 0$, isto é, quando os dois objetos comparados são iguais.

A utilização do modelo espaço de vetores permite que sejam utilizadas diversas equações para calcular o grau de associação entre os criptogramas da coleção de entrada. Essas equações são normalmente aplicadas na determinação da similaridade ou distância entre vetores. Esse cálculo deve ser feito entre cada par de vetores, de maneira que exista um valor de similaridade ou distância para cada par de criptogramas da coleção.

Como cada posição de um vetor representa um bloco, o cálculo é feito passo a passo, onde cada passo representa a aplicação da fórmula considerada (seções 3.1.1.1.1 a 3.1.1.1.9) a um par de blocos, sendo um bloco de cada vetor. Assim, considerando os vetores $\vec{d}_1 = (d_{1,1}, d_{2,1}, ..., d_{n,1})$ e $\vec{d}_2 = (d_{1,2}, d_{2,2}, ..., d_{n,2})$ a cada passo, são calculados $d_{i,1}$ e $d_{i,2}$ por meio da fórmula apropriada.

Para o cálculo das similaridades ou das distâncias, a coleção de objetos será modelada como uma coleção de vetores, os quais representam os objetos da coleção, onde cada coordenada de um vetor representa a quantidade de ocorrências de uma unidade léxica nesse vetor. Assim, uma coordenada representa a freqüência de uma palavra, no caso de documentos ou a freqüência de um bloco, no caso de criptogramas.

Com relação ao cálculo da similaridade ou da distância, é necessário considerar que uma medida em particular pode afetar o resultado do agrupamento (RASMUSSEN, 1992). Além disso, pode-se verificar em Jain (1999) a definição de um algoritmo com cinco passos para realizar um agrupamento, onde um dos passos é a definição de um padrão de medida de similaridade que seja apropriada ao domínio dos dados de entrada. Como não se conhece tal padrão de medida que seja apropriada aos criptogramas, serão utilizadas nove equações para medir a similaridade ou a distância entre os criptogramas da coleção de entrada, com a finalidade de identificar se alguma dessas equações pode produzir melhores resultados de agrupamento, quando comparadas com a medida do ângulo do co-seno utilizada em Carvalho (2006) e assim, defini-la como medida apropriada a uma coleção de entrada composta por criptogramas.

Nas próximas seções, serão apresentadas as medidas de similaridades que serão usadas neste trabalho.

Será considerada a seguinte notação: c_i e c_j par de criptogramas para o qual se deseja obter a similaridade ou a distância. *n* o número de termos sem repetição na coleção de entrada. $c_{i,k}$ o *k-ésimo* termo (bloco) no vetor c_i que representa o criptograma *i*. O vetor c_i é definido como $\vec{c_i} = (c_{1,i}, c_{2,i}, ..., c_{n,i})$. $c_{j,k}$ é o *k-ésimo* termo (bloco) no vetor c_j que representa o criptograma *j*. O vetor c_j é definido como $\vec{c_j} = (c_{1,j}, c_{2,j}, ..., c_{n,j})$.

3.1.1.1.1 COEFICIENTE SIMPLE-MATCHING

O Coeficiente *Simple-matching* considera apenas as dimensões nas quais os vetores possuem valores diferentes de zero (MANNING, 2003). Por ser uma medida de similaridade, quanto maior o valor do resultado do cálculo maior a similaridade entre os objetos. O valor da similaridade é dado pela fórmula 3.2. Intervalo de valores [0,∞[.

$$Similaridade_{Simple-matching}(c_i, c_j) = \sum_{k=1}^{n} (c_{i,k} \times c_{j,k}) \qquad \text{Fórmula 3.2}$$

3.1.1.1.2 COEFICIENTE DICE

Utilizando o Coeficiente *Dice* a similaridade é obtida multiplicando o produto interno de dois vetores por dois e dividindo pela soma das normas desses dois vetores. Por ser uma medida de similaridade, quanto maior o valor do resultado do cálculo maior a similaridade entre os objetos. O valor da similaridade é dado pela fórmula 3.3. Intervalo de valores [0,1].

$$Similaridade_{Dice}(c_i, c_j) = \frac{2 \times \sum_{k=1}^{n} (c_{i,k} \times c_{j,k})}{\sum_{k=1}^{n} (c_{i,k})^2 + \sum_{k=1}^{n} (c_{j,k})^2} \qquad \text{Fórmula 3.3}$$

3.1.1.1.3 MEDIDA DO ÂNGULO DO CO-SENO

É uma medida de similaridade usada para identificar a separação angular de dois vetores (TEKNOMO, 2006), isto é, o co-seno do ângulo entre dois vetores. Esta medida é bastante

parecida com o Coeficiente *Dice*, contudo penaliza[16] menos nos casos em que o número de termos diferentes de zero apresenta uma grande diferença entre os dois vetores (MANNING, 2003). Por ser uma medida de similaridade, quanto maior o valor do resultado do cálculo maior a similaridade entre os objetos. O valor da similaridade é dado pela fórmula 3.4. Intervalo de valores [-1,1]. Neste trabalho, o intervalo de valores utilizados será [0,1].

$$Similaridade_{Co-seno}(c_i, c_j) = \frac{\sum_{k=1}^{n}(c_{i,k} \times c_{j,k})}{\sqrt{\sum_{k=1}^{n}(c_{i,k})^2 \times \sum_{k=1}^{n}(c_{j,k})^2}} \qquad \text{Fórmula 3.4}$$

Pela fórmula 3.4, podemos notar que o valor da similaridade é dado pelo produto escalar dos vetores c_i e c_j dividido pelo produto dos módulos destes vetores. Para objetos textuais, é considerada a medida mais adequada (SALTON, 1987).

3.1.1.1.4 COEFICIENTE JACCARD

Esta medida de similaridade, também conhecida Coeficiente Tanimoto (KOHONEN, 2001), é utilizada comumente na biologia e em química (CHAPMAN, 2006). Esta medida penaliza muito mais um pequeno número de entradas iguais nos vetores, isto é, quando os vetores compartilham poucos termos, do que o Coeficiente *Dice* (MANNING, 2003) e (CHAPMAN, 2006). Por ser uma medida de similaridade, quanto maior o valor do resultado do cálculo maior a similaridade entre os objetos. O valor da similaridade é dado pela fórmula 3.5. Intervalo de valores [0,1].

$$Similaridade_{Jaccard}(c_i, c_j) = \frac{\sum_{k=1}^{n}(c_{i,k} \times c_{j,k})}{\sum_{k=1}^{n}(c_{i,k})^2 + \sum_{k=1}^{n}(c_{j,k})^2 - \sum_{k=1}^{n}(c_{i,k} \times c_{j,k})} \qquad \text{Fórmula 3.5}$$

[16] Neste contexto, penalizar significa reduzir o valor de similaridade, ou seja, quanto mais penaliza menor se torna o valor da similaridade.

3.1.1.1.5 COEFICIENTE OVERLAP

É uma medida de similaridade de uso geral a qual é menos sensível as variações nos tamanhos dos documentos, se comparada à medida do ângulo do co-seno (SULLIVAN, 2006). Se o vetor $c_{i,k}$ é um subconjunto de $c_{j,k}$ ou vice-versa, ou seja, se $c_{i,k} \subseteq c_{j,k}$ ou $c_{j,k} \subseteq c_{i,k}$ então o valor da similaridade é 1 (CHAPMAN, 2006) e (MANNING, 2003). Por ser uma medida de similaridade, quanto maior o valor do resultado do cálculo maior a similaridade entre os objetos. O valor da similaridade é dado pela fórmula 3.6. Intervalo de valores [0,1].

$$Similaridade_{Overlap}(c_i, c_j) = \frac{\sum_{k=1}^{n}(c_{i,k} \times c_{j,k})}{\min(\sum_{k=1}^{n}(c_{i,k})^2, \sum_{k=1}^{n}(c_{j,k})^2)} \qquad \text{Fórmula 3.6}$$

3.1.1.1.6 DISTÂNCIA EUCLIDIANA

É uma medida de dissimilaridade com amplo uso na determinação da distância entre dois vetores (JAIN, 1999) e (TEKNOMO, 2006). Como mede a distância, quanto menor o valor do resultado do cálculo maior a similaridade entre os objetos. O valor da dissimilaridade é dado pela fórmula 3.7. Intervalo de valores [0,∞[.

$$Dissimilaridade_{Euclidiana}(c_i, c_j) = \sqrt{\sum_{k=1}^{n}(c_{i,k} - c_{j,k})^2} \qquad \text{Fórmula 3.7}$$

3.1.1.1.7 DISTÂNCIA MANHATTAN

Também conhecida como *City Block*, é uma medida de dissimilaridade que examina a diferença absoluta entre um par de vetores (TEKNOMO, 2006). Como mede a distância, quanto menor o valor do resultado do cálculo maior a similaridade entre os objetos. O valor da dissimilaridade é dado pela fórmula 3.8. Intervalo de valores [0,∞[.

$$Dissimilaridade_{manhattan}(c_i, c_j) = \sum_{k=1}^{n} \left| c_{i,k} - c_{j,k} \right| \qquad \text{Fórmula 3.8}$$

3.1.1.1.8 DISTÂNCIA CANBERRA

A Distância Canberra utiliza o somatório de uma série de frações para determinar a dissimilaridade entre dois vetores. Esta distância tende a efetuar mudanças sensíveis no resultado do cálculo da similaridade, quando ambas as coordenadas utilizadas em uma fração possuem valores próximos de 0 (TEKNOMO, 2006). Isto quer dizer que, se tivermos apenas um dos resultados do cálculo de uma fração próximo de 0, o valor da distância será reduzido consideravelmente. Como mede a distância, quanto menor o valor do resultado do cálculo maior a similaridade entre os objetos. O valor da dissimilaridade é dado pela fórmula 3.9. Intervalo de valores [0,∞[.

$$Dissimilaridade_{Canberra}(c_i, c_j) = \sum_{k=1}^{n} \frac{\left| c_{i,k} - c_{j,k} \right|}{\left| c_{i,k} \right| + \left| c_{j,k} \right|} \qquad \text{Fórmula 3.9}$$

3.1.1.1.9 DISTÂNCIA BRAY-CURTIS

Esta distância, também chamada de distância Sorensen, é um método normalizado de uso comum em Botânica e Ecologia, onde a sua normalização é obtida pela divisão do valor do somatório da diferença absoluta pelo somatório das coordenadas (TEKNOMO, 2006). Uma vez que obtém a distância, quanto menor o valor do resultado do cálculo maior a similaridade entre os objetos. O valor da dissimilaridade é dado pela fórmula 3.10. Como neste trabalho serão considerados apenas valores positivos, o valor da distância com nessa medida estará sempre no intervalo [0,1].

$$Dissimilaridade_{Bray-Curtis}(c_i, c_j) = \frac{\sum_{k=1}^{n} \left| c_{i,k} - c_{j,k} \right|}{\sum_{k=1}^{n} (c_{i,k} + c_{j,k})} \qquad \text{Fórmula 3.10}$$

3.1.1.2 CONSIDERAÇÃO SOBRE A AUSÊNCIA E PRESENÇA DE BLOCOS NOS VETORES

Deve ser observada uma característica importante das medidas de similaridade e distância adotadas neste trabalho. Esta característica diz respeito à maneira como os passos das equações são realizados, de acordo com a ausência ou presença de um determinado bloco nos vetores. Assim, podem-se ter quatro situações possíveis (A, B, C e D), conforme a tabela 3.1.

TAB. 3.1 – Ausência e presença de blocos (adaptado de MEYER, 2002)

		Vetor 2	
		Presente	Ausente
Vetor 1	Presente	A	B
	Ausente	C	D

Como foi dito anteriormente, a posição do vetor relativa ao bloco recebe o valor um, se o

pertencentes à coleção de entrada podem ser pré-calculados, o que evita o cálculo da similaridade no momento da execução do agrupamento, simplificando este processo (JAIN, 1999). A matriz será usa na tarefa de agrupamento (seção 3.1.5). Assim, haverá uma matriz de similaridades ou distâncias com os valores de similaridades ou distância para cada par de criptogramas.

Na tabela 3.2 e na figura 3.2, podem ser vistos exemplos de matriz de similaridades.

TAB. 3.2 – Matriz de similaridades (CARVALHO, 2006)

Documentos	1	2	3	4
1	1	0,183	0	0,365
2	0,183	1	0	0,4
3	0	0	1	0
4	0,365	0,4	0	1

Nota-se que se trata de uma matriz simétrica, isto é, $S_{i,j} = S_{j,i}$. Assim, é suficiente usar a matriz triangular inferior ou superior. A diagonal deve ser desconsiderada também, já que representa a similaridade entre o documento e ele mesmo, a qual é sempre igual a 1.

$$S = \begin{vmatrix} S_{21} & & & & \\ S_{31} & S_{32} & & & \\ S_{41} & S_{42} & S_{43} & & \\ \vdots & \vdots & \vdots & \ddots & \\ S_{N1} & S_{N2} & S_{N3} & \cdots & S_{N(N-1)} \end{vmatrix}$$

FIG. 3.2: Matriz de similaridades (RASMUSSEN, 1992)

3.1.2 ARQUIVO INVERTIDO

Conforme visto na seção 3.1, um arquivo invertido é uma estrutura adequada para Sistemas RI, sendo um tipo de arquivo indexado (FRAKES, 1992). O processo de transformação de uma coleção de documentos em um arquivo invertido é chamado *indexação* (CARVALHO, 2006).

Essa estrutura será utilizada neste trabalho para o armazenamento da coleção de criptogramas. A seguir será mostrado como esse tipo de arquivo é usado pelos sistemas RI e como pode ser usado no caso dos criptogramas.

Um arquivo invertido pode ser estruturado em três campos: uma palavra-chave ou índice, relativa a uma palavra da coleção de objetos, um apontador, o qual indica a que objeto a palavra-chave pertence (HARMAN, 1992a) e um campo indicando a freqüência da palavra chave no objeto (CARVALHO, 2006) e (FRAKES, 1992). Na tabela 3.3, podemos ver um exemplo de arquivo invertido para os documentos da seção 3.1.1.

TAB. 3.3 – Arquivo invertido para os documentos da seção 3.1.1

Palavra-chave	**Doc.**	**Freq.**	**Doc.**	**Freq.**	**Doc.**	**Freq.**
Recuperação	1	1			3	1
Informação	1	1	2	1		
Criptografia	1	1	2	1	3	1

Verificando a tabela 3.3 pode-se perceber que teremos uma coluna documento e uma coluna freqüência para cada documento da coleção. Vê-se também que, diferente do modelo vetorial, a ausência da palavra não gera a freqüência zero. Simplesmente o espaço fica sem preenchimento. No exemplo apresentado, existe apenas palavras com freqüência igual a 1. Em uma coleção de objetos maior, é possível ter freqüências maiores.

Para os criptogramas, uma palavra chave é um bloco, o documento é o criptograma e a freqüência e a quantidade de vezes que um bloco aparece em um criptograma.

3.1.3 LEMATIZAÇÃO (STEMMING), REMOÇÃO DE *STOPLIST* E PESAGEM (*WEIGHTING*)

Estas operações têm a finalidade de melhorar o processo de recuperação de informações, sendo aplicadas a sistemas específicos, após uma análise sobre a adequação. A lematização e a remoção de *stoplist* têm ainda a finalidade de reduzir o tamanho dos arquivos de índices[17]

[17] Neste trabalho, os arquivos invertidos.

(FRAKES, 1992b) e (FOX, 1992). Desta maneira, serão apresentados nesta seção os detalhes destas operações. Na Seção 3.1.4 é apresentado um exemplo de uso destas operações, onde pode ser vista a justificativa por que neste trabalho somente a pesagem será utilizada.

A lematização, também conhecida como *word stemming*, tem a finalidade de reduzir as palavras aos seus radicais. Assim, pode-se utilizar em uma consulta uma palavra genérica, como: criptografia, e recuperar documentos que possuam as palavras: criptologia, criptoanálise, criptoanalista, criptográfico e assim por diante, após submetermos todas essas palavras à operação de lematização (figura 3.3).

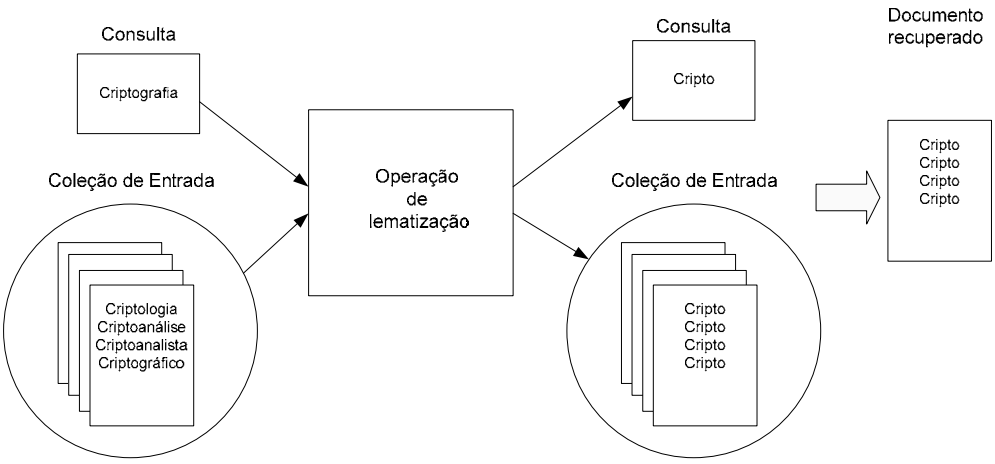


FIG. 3.3: Uma operação de consulta com o uso de lematização

As desvantagens dessa operação estão relacionadas a uma redução muito grande ou muito pequena da palavra, o que pode ocasionar o relacionamento de termos genéricos aos radicais de outras palavras, no primeiro caso, e o não relacionamento de termos genéricos ao radical da própria palavra, no segundo caso.

A remoção de *stoplist* consiste na remoção de palavras pouco expressivas, conhecidas como *stopwords* (FOX, 1992) da coleção de documentos. Desta forma, é preciso montar uma lista (*stoplist*) com palavras que freqüentemente ocorrem no idioma considerado, tais como preposições, artigos e partículas de texto. Palavras com sentido muito genérico, no contexto da coleção tratada, também devem ser incluídas na *stoplist*. Por questão de eficiência, esta operação deve ser realizada antes de se incluir as palavras no arquivo de índice. Assim, antes

de realizar a indexação, deve-se consultar a *stoplist*. Caso a palavra considera esteja na *stoplist*, ela não deve ser incluída no arquivo de índice.

Pode-se incluir na *stoplist*, palavras com dois ou menos caracteres, dado a pobreza do valor semântico dessas palavras. Além disso, é recomendável realizar um *filtro de caracteres*, retirando números, fórmulas matemáticas, símbolos, caracteres especiais, pontuação e acentuação (XEXÉO, 2006).

A pesagem (*weighting*) é a associação de um valor numérico, denominado peso, aos termos de uma coleção, de acordo com a relevância que cada um desses termos possui na coleção. É necessário, então, definir essa relevância. Uma abordagem para a atribuição de pesos é a distribuição estatística dos termos nos documentos da coleção (FRAKES, 1992). Isto é, verifica-se a freqüência dos termos em cada documento (HARMAN), ao invés de simplesmente identificar se o termo existe ou não em um documento.

Baseados na definição apresentada acima, nós teríamos uma nova descrição formal para o modelo vetorial: sejam d_1 e d_2, dois documentos. O valor relacionado à $d_{i,1}$, onde *i* representa a *i-ésima* palavra do documento 1, é a freqüência do termo *i* no documento 1, se o documento 1 contém a palavra, ou 0 caso contrário. O valor relacionado à $d_{i,2}$, onde *i* representa a *i-ésima* palavra do documento 2, é a freqüência do termo *i* no documento 2, se o documento 2 contém a palavra, ou 0 caso contrário. Então, o vetor para o documento 1 é definido como $\vec{d}_1 = (d_{1,1}, d_{2,1}, ..., d_{n,1})$, assim como o vetor para o documento 2 é representado por $\vec{d}_2 = (d_{1,2}, d_{2,2}, ..., d_{n,2})$, onde *n* é o número de palavras únicas na coleção de documentos. A coleção de documentos é denotada por $D = (d_1, d_2, ..., d_n)$.

3.1.4 TRATAMENTO DE UMA COLEÇÃO DE OBJETOS

Com base na figura 3.4, será apresentado um exemplo do tratamento de uma coleção de objetos, por meio da indexação, a qual envolverá as operações de remoção de *stoplist*, filtro de caracteres, e pesagem (figura 3.5).

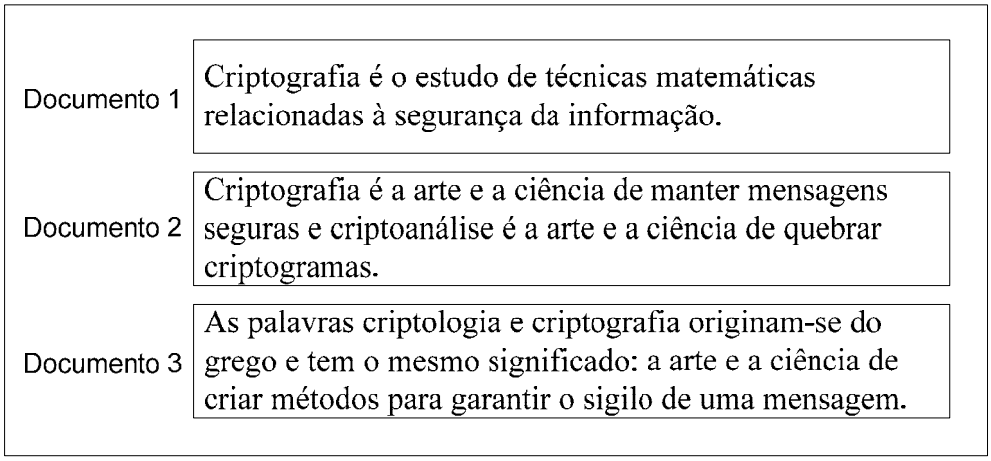


FIG. 3.4: Coleção de documentos com três objetos

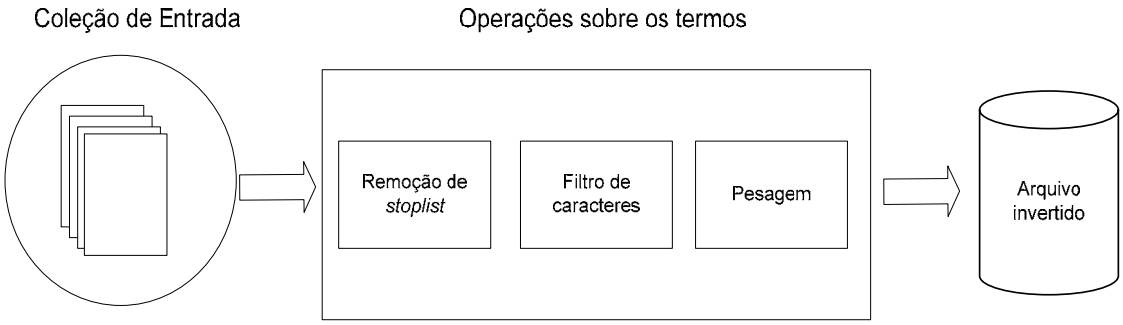


FIG. 3.5: O processo de indexação

A figura 3.6 mostra como ficam os documentos da coleção após a realização das operações sobre os termos.

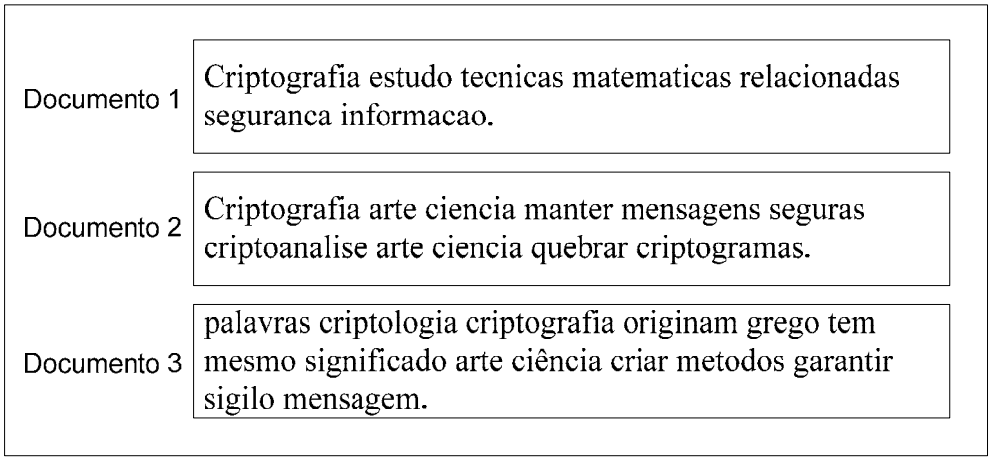


FIG. 3.6: Coleção de documentos com três objetos, após as operações sobre os termos

Na tabela 3.4, pode-se verificar o arquivo indexado criado como resultado do processo de indexação. Percebe-se pelas entradas da tabela, na coluna palavra-chave, o resultado das operações realizadas, a qual diminuiu o número de entradas e modificou a ortografia das palavras, decorrente da aplicação do filtro de caracteres. Nota-se que uma palavra no plural gera um novo índice. O uso da lematização poderia eliminar entradas deste tipo, transformando “mensagem” e “mensagens”, assim como “segurança” e “seguras”, em um só índice. Além disso, a palavra “criptografia” aparece em todos os documentos da coleção, logo não é uma boa palavra para distinguir ou aproximar documentos em particular, assim deve ser incluída na *stoplist*.

Na tabela 3.5, está a representação do espaço vetorial após o resultado da indexação. Neste exemplo, após as operações realizadas haveria um total de 27 termos únicos (n = 27), logo o espaço vetorial terá 27 dimensões.

Observando ambas as tabelas (3.4 e 3.5), verifica-se que as mesmas possuem uma estrutura bem parecida. Pode-se considerar, então, que o arquivo invertido é a implementação do modelo de espaço de vetores, embora, na prática, seja possível implementar o modelo de espaço de vetores sem a necessidade do arquivo invertido.

TAB. 3.4 – Arquivo invertido criado após a indexação

Palavra-chave	**Doc.**	**Freq.**	**Doc.**	**Freq.**	**Doc.**	**Freq.**
criptografia	1	1	2	1	3	1
estudo	1	1				
tecnicas	1	1				
matematicas	1	1				
relacionadas	1	1				
seguranca	1	1				
informacao	1	1				
arte	2	2	3	1		
ciencia	2	2	3	1		
manter	2	1				
mensagens	2	1				
seguras	2	1				
criptoanalise	2	1				
quebrar	2	1				
criptogramas	2	1				
palavras	3	1				
criptologia	3	1				
originam	3	1				
grego	3	1				
tem	3	1				
mesmo	3	1				
significado	3	1				
criar	3	1				
metodos	3	1				
garantir	3	1				
sigilo	3	1				
mensagem	3	1				

As palavras, no contexto dos criptogramas, são grupos de unidades léxicas de um alfabeto binário, as quais têm tamanho único e dependente do algoritmo criptográfico utilizado para cifrá-las. Disto, é possível chamar esses grupos léxicos de *criptotermos*. Assim, pode-se ver que não há sentido em usar as operações de lematização, remoção de *stoplist* e filtro de caracteres. A única ressalva quanto à remoção de *stoplist*, seria a ocorrência de um ou mais criptotermos em todos os criptogramas da coleção, os quais poderiam ser removidos. Como a operação de pesagem pode ser aplicada aos criptogramas, ela será utilizada neste trabalho.

TAB. 3.5 – Espaço vetorial após a indexação

n	Palavra	Doc. 1	Doc. 2	Doc. 3
1	criptografia	1	1	1
2	Estudo	1	0	0
3	Técnicas	1	0	0
4	matematicas	1	0	0
5	relacionadas	1	0	0
6	seguranca	1	0	0
7	informacao	1	0	0
8	arte	0	2	1
9	ciencia	0	2	1
10	manter	0	1	0
11	mensagens	0	1	0
12	seguras	0	1	0
13	criptoanalise	0	1	0
14	quebrar	0	1	0
15	criptogramas	0	1	0
16	palavras	0	0	1
17	criptologia	0	0	1
18	originam	0	0	1
19	grego	0	0	1
20	tem	0	0	1
21	mesmo	0	0	1
22	significado	0	0	1

3.1.5 AGRUPAMENTO (*CLUSTERING*)

O agrupamento (*clustering*) é uma tarefa na qual, dada uma coleção *C* de objetos, consiste em dividir os *n* objetos desta coleção em *m* grupos. O valor de *m* é normalmente desconhecido, o que caracteriza o modo não supervisionado do processo de agrupamento (JAIN, 1999). O agrupamento pode ser usado como um fim em si próprio ou com a finalidade de melhorar o resultado da aplicação de sistema RI (RASMUSSEN, 1992).

O agrupamento pode ser aplicado a diversos tipos de objetos, como: imagens, padrões, palavras, documentos e outros (WANNER, 2005), e é a base para muitas aplicações (JAIN, 1999), como *Data mining* (GOLDSCHMIDT, 2005), *Text mining* e lingüística computacional (WANNER, 2005).

Uma das tarefas na qual se pode utilizar o agrupamento seria para agrupar documentos por assuntos (figura 3.7). Neste caso, é necessário extrair características que identifique cada um dos documentos. Assim, o conjunto de documentos que possuir as mesmas características formaria um grupo. Pode-se restringir essas características a uma única característica, a qual utilizaria uma abordagem estatística baseada na freqüência das palavras dos documentos. Está característica seria a *similaridade* entre os documentos. Desta forma, quanto mais palavras em comum um documento tiver com outro, maior a chance destes dois documentos tratarem do mesmo assunto e, assim, pertencerem ao mesmo grupo.

O resultado do agrupamento são grupos nos quais os documentos de um grupo possuem alta similaridade entre si e alta dissimilaridade com os documentos de outro grupo (JAIN, 1999), (WANNER, 2005) e (RASMUSSEN, 1992).

Considerando os criptogramas (figura 3.8), tem-se que a aplicação de uma chave qualquer a um algoritmo criptográfico determina uma linguagem particular e dá origem a criptogramas escritos nesta linguagem (CARVALHO, 2006). Assim criptogramas produzidos com uma mesma chave serão mais similares, uma vez que compartilharão o mesmo conjunto de elementos léxicos. Logo, o agrupamento aplicado aos criptogramas tem como objetivo separar a coleção de entrada em grupos onde os criptogramas que compõem tal grupo tenham sido cifrados com a mesma chave e o mesmo algoritmo criptográfico, conforme a figura 3.8.

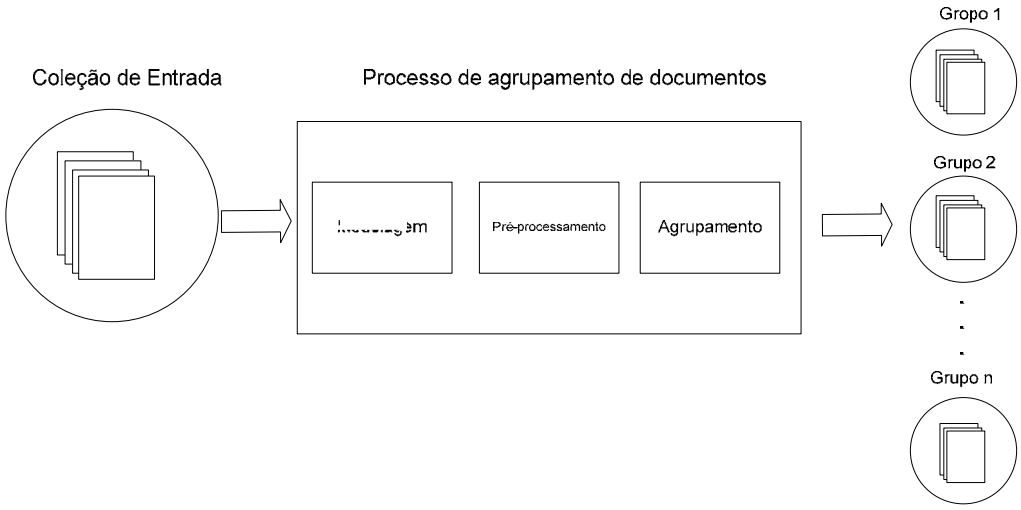


FIG. 3.7: O processo de agrupamento

As técnicas de agrupamento são classificadas de acordo com o tipo de estrutura que geram e podem ser hierárquicos ou particionais (não-hierárquicos) (JAIN, 1999) e (RASMUSSEN, 1992). O agrupamento hierárquico é usado quando o conjunto de objetos de entrada não tem uma expectativa de divisão em grupos bem definidos. Os dois principais tipos de algoritmos para agrupamento hierárquico atualmente em uso são: algoritmos divisivos, os quais a partir de um único grupo contendo todos os objetos da coleção, particionam recursivamente cada cluster até que algum critério de parada seja alcançado (WANNER, 2005); e algoritmos aglomerativos, os quais iniciam com um grupo para cada documento e prosseguem juntando estes grupos par a par até que um critério de parada seja alcançado.

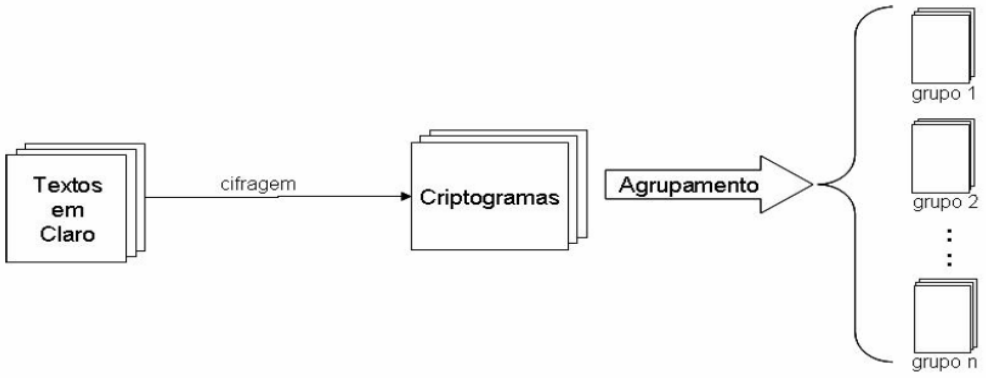


FIG. 3.8: Processo de agrupamento de criptogramas (CARVALHO, 2006)

No agrupamento particional a coleção *C* se torna uma simples divisão em *m* grupos. Esta técnica é mais bem usada para aplicações com grandes conjuntos de dados (JAIN, 1999).

3.1.5.1 MÉTODOS HIERÁRQUICOS

Os métodos hierárquicos são os mais utilizados para o agrupamento de documentos (RASMUSSEN, 1992). Um algoritmo hierárquico aglomerativo, pode ser visto em Wanner (2005) e se constitui dos seguintes passos:

a) Iniciar os grupos, assumindo que cada documento pertencerá a um único grupo;

b) Calcular a matriz de similaridades entre todos os pares de grupos;

c) Juntar o par de grupos mais similar;

d) Atualizar a matriz de similaridades, de maneira que o novo grupo possa ser comparado com os grupos restantes;

e) Repetir os passos *c* e *d* até que um critério de parada seja alcançado.

Como resultado da aplicação desse algoritmo, um gráfico denominado dendograma (figura 3.9) é formado, o qual demonstra a estrutura final do agrupamento, representando a ordem em que as junções dos grupos ocorreram (RASMUSSEN, 1992) e (JAIN, 1999). Observando a figura 3.9, podemos notar que os grupos foram formados a partir de um determinado valor de similaridade.

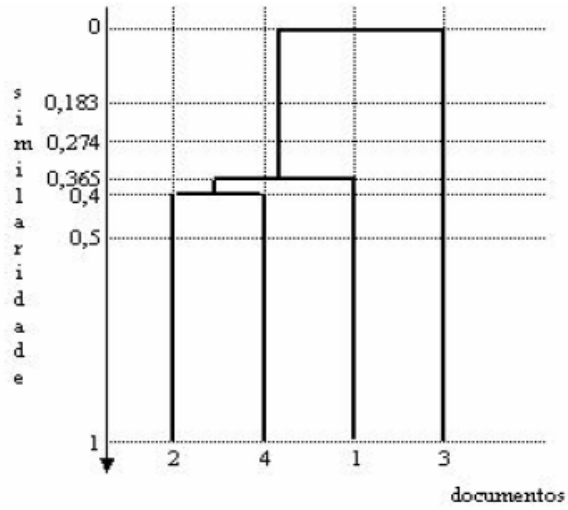


FIG. 3.9: Dendograma (CARVALHO, 2006)

Existem diferentes métodos para definir a forma de dividir ou juntar os grupos, como: *single-link, complete-link, group average-link* ou método de Ward (RASMUSSEN, 1992).

3.1.5.1.1 LIGAÇÃO SIMPLES (*SINGLE- LINK*)

Neste método, a cada passo, o par mais similar de objetos que ainda não está no mesmo grupo é fundido em um grupo.

O método *single-link* possui a desvantagem de ser influenciado por um efeito em cadeia. Isto quer dizer que padrões ruidosos entre grupos de pontos esféricos, induzem este algoritmo à formação de grupos dispersos ou alongados (JAIN, 1999) (figura 3.10). A dispersão é um item particularmente importante, uma vez que permite que dois objetos quaisquer que estejam em grupo, possuam valor de similaridade mais baixo que a similaridade do próprio grupo, o que é apropriado para o uso com criptogramas (CARVALHO, 2006). Em contrapartida, tem a capacidade de extrair grupos em pontos agrupados de maneira concêntrica (figura 3.11).

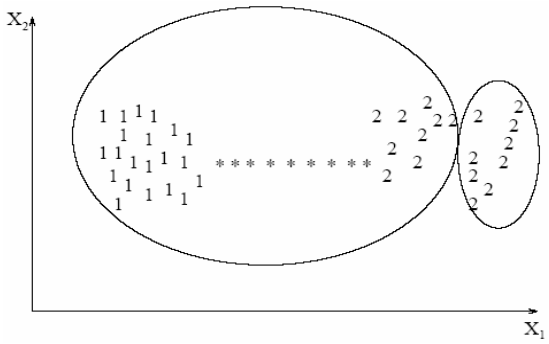


FIG. 3.10: Formação de grupos alongados influenciada por ruído (JAIN, 1999)

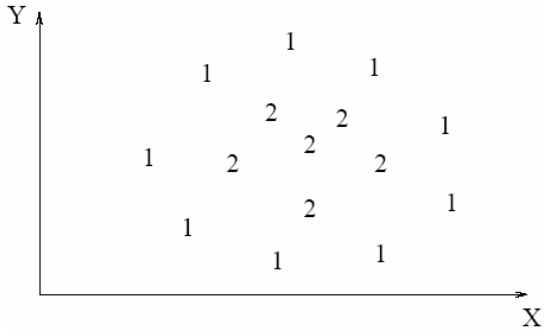


FIG. 3.11: Dois grupos de pontos concêntricos no plano (JAIN, 1999)

3.1.5.1.2 LIGAÇÃO COMPLETA (*COMPLETE-LINK*)

Neste método, a cada passo, o par menos similar de objetos que ainda não está no mesmo grupo é fundido em um grupo.

Este método produz grupos mais compactos, produzindo hierarquias mais úteis do que o *single-link* para muitas aplicações (JAIN, 1999). Diferente do *single-link*, produz grupos adequados em pontos agrupados de maneira esférica, isto quer dizer que os ruídos têm pouca ou nenhuma influência neste algoritmo para este caso (figura 3.12). Entretanto, o *complete-link* não consegue extrair grupos em pontos agrupados de maneira concêntrica (JAIN, 1999).

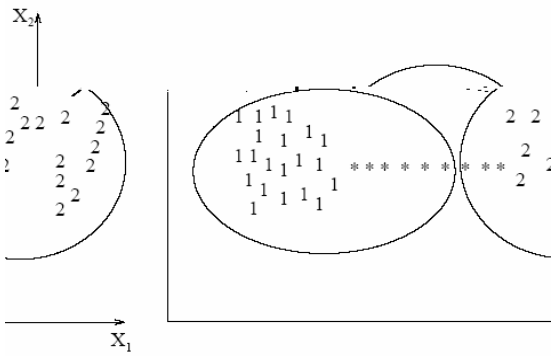


FIG. 3.12: Formação de grupos adequados sem a influência de ruído (JAIN, 1999)

3.1.5.1.3 LIGAÇÃO POR MÉDIA DOS GRUPOS (*GROUP AVERAGE-LINK*)

Este método utiliza a média ponderada das similaridades dos grupos para fazer a junção de dois grupos. Neste caso, todos os objetos dos grupos anteriormente unidos contribuem para o cálculo da nova similaridade. A ligação pela média dos grupos se constitui em uma estrutura intermediária entre a ligação simples e a ligação completa (RASMUSSEN, 1992).

A partir da matriz de similaridades, uma nova similaridade é calculada, utilizando-se a formúla 3.11 (XEXÉO, 2006).

$$Similaridade(g_{ij}, g_k) = \frac{m_i \times Sim(g_i, g_k) + m_j \times Sim(g_j, g_k)}{m_i + m_j} \qquad \text{Fórmula 3.11}$$

Onde,

g_i, g_j são dois grupos quaisquer.

m_i, m_j número de itens nos grupos g_i, g_j, respectivamente.

g_{ij} grupo resultante da fusão dos clusters g_i e g_j.

g_k grupo para o qual pretende-se obter a similaridade com o grupo g_{ij}.

3.1.6 MÉTODOS DE AVALIAÇÃO

Neste trabalho, utilizou-se cinco medidas de similaridade: Dice, Ângulo do Co-seno, Jaccard, Overlap e Simple-matching; e quatro medidas de dissimilaridade (distâncias): Euclidiana, Manhattan, Canberra e Bray-Curtis. Será utilizado o Coeficiente de Correlação de Pearson para avaliação das medidas de similaridade e distância. Porém, antes de se empregar os métodos estatísticos nas medidas, é necessário converter os valores das respectivas matrizes em valor de dissimilaridade, para que todas as medidas demonstrem o grau de associação entre os criptogramas de maneira diretamente proporcional.

Os valores foram convertidos como segue: sejam *i* e *j* dois criptogramas de uma coleção de entrada e seja $s_{i,j}$ o valor da similaridade entre esses dois objetos. Então, o valor da dissimilaridade $d_{i,j}$ entre esses dois objetos é dado pela fórmula 3.12.

$$d_{i,j} = 1 - s_{i,j} \qquad \text{Fórmula 3.12}$$

A decisão de converter os valores das medidas de similaridade em valores de dissimilaridade e não o contrário, se deveu ao fato de que as todas as medidas de similaridade, utilizadas neste trabalho, possuírem valores no intervalo [0,1], exceto no caso da medida Simple-matching, cujos valores estão no intervalo [0,∞[. Já para as medidas de distância, estas possuem valores no intervalo [0,∞[, exceto no caso da distancia Bray-Curtis, cujos valores estão no intervalo [0,1]. Desta forma, converter as medidas de distância seria inadequado já que os valores para as essas medidas podem ser bem altos (ver apêndice 2), se comparados aos valores de similaridade.

Nesta avaliação, não foi considerada a medida *Simple-matching*, já que a mesma possui intervalo de valores semelhantes às medidas distância, mas indica associação inversamente proporcional a estas medidas. Além disso, não é possível converter os seus valores pela fórmula 3.12, pois resultaria em um número negativo.

3.1.6.1 MÉTODO ESTATÍSTICO PARA AVALIAÇÃO DAS MEDIDAS DE SIMILARIDADE E DISTÂNCIA

No primeiro passo para identificação de como as medidas de similaridade e distância afetam o resultado do agrupamento, é preciso verificar qual o grau de relacionamento entre essas medidas. Para essa finalidade será utilizado o Coeficiente de Correlação Momento-Produto ou Coeficiente Linear de Pearson, denotado por *r*. Assim, a partir das matrizes de similaridades e distâncias geradas pelos coeficientes descritos anteriormente, utilizamos a fórmula 3.13 (IEMMA, 1992) para obtenção do valor de *r*.

$$r = \frac{\sum_{i=1}^{n-1}\sum_{j=2}^{n}(x_{ij}-\bar{x})(y_{ij}-\bar{y})}{\sqrt{[\sum_{i=1}^{n-1}\sum_{j=2}^{n}(x_{ij}-\bar{x})^2][\sum_{i=1}^{n-1}\sum_{j=2}^{n}(y_{ij}-\bar{y})^2]}} \qquad \text{Fórmula 3.13}$$

Onde,

n é o número de criptogramas na coleção de entrada.

x_{ij} é o valor da similaridade para os criptogramas *i* e *j*, da matriz de similaridade produzida pelo coeficiente *x*.

y_{ij} é o valor da similaridade para os criptogramas *i* e *j*, da matriz de similaridade produzida pelo coeficiente *y*.

$\bar{x}$ é obtido por meio da fórmula 3.14.

$\bar{y}$ é obtido por meio da fórmula 3.15.

.

$$\bar{x} = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=2}^{n}x_{ij} \qquad \text{Fórmula 3.14}$$

$$\bar{y} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=2}^{n} y_{ij}$$ Fórmula 3.15

O cálculo de r é realizado para cada par de matrizes de similaridades, dando origem a uma matriz contendo todos os valores de r para cada par de medidas.

O coeficiente de correlação pode ser considerado uma covariância normalizada. Assim, qualquer par de valores calculados fica restrito ao intervalo [-1,1] (BARROS NETO, 2003). Essa propriedade é bastante apropriada para o caso das medidas sendo utilizadas, uma vez que a distâncias Euclidiana, Manhattan e Canberra tendem ao infinito, podendo possuir valores muito diferentes, quando comparados às outras medidas.

O valor de $r = 0$ indica que não há uma relação linear entre as medidas, entretanto isso não assegura que não haja quaisquer outros tipos de dependência. Quando $r = 1$, então existe uma relação linear perfeita, indicando que quando uma medida aumenta a outra também aumenta. Quando $r = -1$, também existe uma relação linear perfeita, indicando que quando uma medida aumenta a outra diminui (BARROS NETO, 2003).

3.1.6.2 AVALIAÇÃO DO AGRUPAMENTO

Após o processo de agrupamento, é necessário verificar se os grupos produzidos representam os melhores grupos possíveis de serem gerados. A questão principal é definir o que é um bom grupo. Qualquer algoritmo de agrupamento produzirá grupos como resultado de sua execução, independente de a coleção de entrada possuir grupos bem definidos ou não (JAIN, 1999).

Os grupos produzidos são válidos se eles não foram produzidos por acaso ou como um artefato[18] de um algoritmo de agrupamento (JAIN, 1999).

Neste trabalho, serão utilizadas as medidas relativas à *recall* e a *precision* (YATES, 1999) e (FUNG, 2003), demonstrada na figura 3.13, uma vez que estas foram as medidas utilizadas em Carvalho (2006) e pretendemos comparar os resultados obtidos naquele trabalho com os resultados desta dissertação. O Intervalo de valores de ambas as medidas é [0,1], sendo um (1) o melhor resultado possível.

[18] Observação ilusória durante uma medição ou experiência científica e que se deve a imperfeições no método ou na aparelhagem (FERREIRA, 2004).

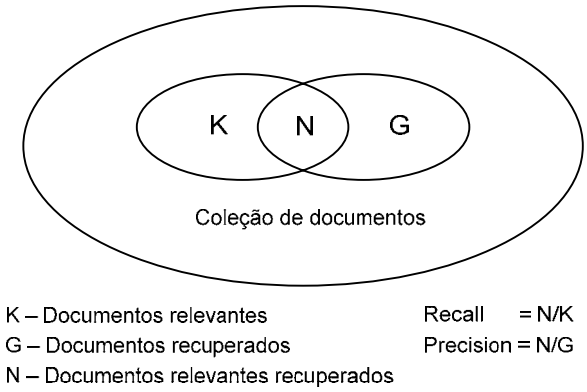


FIG. 3.13: Precision e recall

Suponha que o conjunto k_i representa o grupo completo dos criptogramas que foram cifrados com uma determinada chave. Então, como resultado do processo de agrupamento, o melhor resultado para a *recall* seria recuperar um grupo g_i com todos os elementos de k_i e o melhor resultado para *precision* seria ter neste grupo g_i somente elementos de k_i. Logo, g_i é o conjunto dos criptogramas recuperados. Seja $|k_i|$ o número de elementos no conjunto k_i. Seja $|g_i|$ o número de elementos no grupo g_i. Seja n_i o número de elementos do conjunto k_i presentes no grupo g_i. Então, os valores de *Recall* e *Precision* são obtidos pelas fórmulas 3.16 e 3.17.

$$recall = \frac{n_i}{|k_i|} \qquad \text{Fórmula 3.16}$$

$$precision = \frac{n_i}{|g_i|} \qquad \text{Fórmula 3.17}$$

Deve-se notar que as fórmulas acima determinam a *Recall* e a *Precision* somente para um grupo. Assim, para consolidar os valores dos *m* grupos formados, pode ser utilizado a micro-média (YANG, 1999), que consiste no cálculo dos valores globais para essas duas medidas (fórmulas 3.18, 3.19, 3.20, 3.21 e 3.22).

$$N = \sum_{i=1}^{m} n_i \qquad \text{Fórmula 3.18}$$

$$K = \sum_{i=1}^{m} k_i$$ Fórmula 3.19

$$G = \sum_{i=1}^{m} g_i$$ Fórmula 3.20

$$recall = \frac{N}{K}$$ Fórmula 3.21

$$precision = \frac{N}{G}$$ Fórmula 3.22

De maneira alternativa, poderia ser feito o cálculo por meio da macro-média (YANG, 1999), onde a *recall* e a *Precision* são calculadas localmente, para cada um dos *m* grupos, sendo obtida a média ao final (fórmulas 3.23e 3.24). Inicialmente será utilizada a micro-média no presente trabalho. Mas é indiferente o uso de qualquer uma das formas para a consolidação dos valores, no contexto desta dissertação.

$$recall = \frac{1}{m} \sum_{i=1}^{m} \frac{n_i}{|k_i|}$$ Fórmula 3.23

$$precision = \frac{1}{m} \sum_{i=1}^{m} \frac{n_i}{|g_i|}$$ Fórmula 3.24

4 REDE NEURAL AUTO-ORGANIZÁVEL BASEADA NO MAPA DE KOHONEN APLICADA A CRIPTOGRAMAS

Uma rede neural auto-organizável é um modelo matemático que possui características adequadas para realização de agrupamento de padrões. Ou seja, a partir das entradas fornecidas, um mapa topográfico de características vai se formando no decorrer das iterações (FAUSSET, 1994). A rede é formada por um conjunto de neurônios. Cada neurônio possui um vetor de pesos sinápticos de *n-dimensões* associado a ele. As localizações espaciais destes neurônios no mapa topográfico indicam características contidas nas entradas apresentadas à rede (HAYKIN, 2001). O resultado apresentado na saída é um mapa no qual podem ser visualizados os grupos formados. Neste trabalha será utilizado um mapa bidimensional.

Assim, um Mapa de Kohonen é um tipo de rede neural auto-organizável, baseado em aprendizado competitivo, com treinamento não-supervisionado (HAYKIN, 2001), isto é, quando não conhecemos as classes a priori e apesar disso, ao final da classificação obtemos um conjunto finito de classes (HONKELA[A], 1997) e (KOHONEN, 2001), onde dados abstratos são submetidos à entrada da rede (KOHONEN, 1989), a qual tem entre uma das suas principais aplicações a tarefa de agrupamento (GOLDSCHMIDT, 2005). No mapa resultante, as entradas similares são mapeadas próximas uma das outras (HONKELA[A], 1997).

As entradas podem ser também chamadas de padrões no contexto dos Mapas de Kohonen. Neste trabalho, uma entrada ou um padrão é um criptograma.

Para a formação do mapa, são utilizados três processos: processo competitivo, cooperativo e adaptativo.

4.1 O PROCESSO COMPETITIVO

No processo competitivo, um padrão do conjunto de entradas é escolhido aleatoriamente e é apresentado à rede. Para cada neurônio da rede é calculado um valor, tomando por base o padrão de entrada e o vetor de pesos sinápticos do neurônio. Neste trabalho, o cálculo será realizado por meio da distância Euclidiana (Fórmula 4.1) ou pela medida do ângulo do co-seno (Fórmula 4.2). Contudo, outras medidas podem ser usadas neste cálculo, como as apresentadas na seção 3.1.1.1. Uma vez que o cálculo é feito para todos os neurônios, temos a

idéia de competição, onde o neurônio que obtiver o menor valor, no caso da utilização da distância Euclidiana, será o vencedor. O menor valor quer dizer que o neurônio vencedor possui a menor dissimilaridade com o padrão de entrada, isto é, está mais apto a representar o padrão. No caso da medida do ângulo do co-seno deve ser considerado o maior valor. O neurônio vencedor terá, então, o seu vetor de pesos sinápticos atualizados.

$$Dissimilaridade_{Euclidiana}(c_i, ps_j) = \sqrt{\sum_{k=1}^{n}(c_{i,k} - ps_{j,k})^2} \qquad \text{Fórmula 4.1}$$

$$Similaridade_{Co-seno}(c_i, ps_j) = \frac{\sum_{k=1}^{n}(c_{i,k} \times ps_{j,k})}{\sqrt{\sum_{k=1}^{n}(c_{i,k})^2 \times \sum_{k=1}^{n}(ps_{j,k})^2}} \qquad \text{Fórmula 4.2}$$

Onde $c_{i,k}$ representa o criptotermo k do vetor (ou criptograma) de entrada i, e $ps_{j,k}$ representa o segmento k do vetor de pesos sinápticos relativo ao neurônio j.

4.2 O PROCESSO COOPERATIVO

O neurônio vencedor é o centro de uma vizinhança topológica (HAYKIN, 2001). Isto quer dizer que não só o vencedor terá o seu vetor de pesos sinápticos atualizados, mas também todos os neurônios da sua vizinhança.

O valor de atualização para os neurônios vizinhos deve ir diminuindo à medida que a distância do neurônio vizinho aumenta em relação ao neurônio vencedor. Assim, no processo cooperativo ocorre o cálculo dos valores para a vizinhança do vencedor.

A função de vizinhança é dada pela fórmula 4.3.

$$h_{j,i(d)}(n) = \exp(-\frac{d_{j,i}^2}{2\sigma^2(n)}) \qquad \text{Fórmula 4.3}$$

Onde $d_{j,i}^2$ é a distância entre o neurônio vencedor i e um neurônio vizinho j. n é um intervalo de tempo discreto, o qual pode ser representado pela iteração corrente em um determinando momento do tempo. O σ é dado pela fórmula 4.4.

$$\sigma(n) = \sigma_0 \exp(-\frac{n}{\tau_1})$$ Fórmula 4.4

Onde σ_0 é o tamanho inicial da vizinhança. n é um intervalo de tempo discreto, o qual pode ser representado pela iteração corrente em um determinando momento do tempo e τ_1 é uma constante de tempo dada pela fórmula 4.5.

$$\tau_1 = \frac{N}{\log \sigma_0}$$ Fórmula 4.5

Onde N é o número total de iterações.

4.3 O PROCESSO ADAPTATIVO

No processo adaptativo ocorre o aprendizado da rede. Baseado nos resultados dos processos anteriores, os valores para a atualização dos vetores de pesos sinápticos do vencedor e dos seus vizinhos são calculados, de acordo com a fórmula 4.6. Caso um neurônio não pertença à vizinhança h_i, ou seja, o conjunto de todos os vizinhos do neurônio vencedor i, o seu vetor de pesos sinápticos permanecerá o mesmo.

$$ps_j(n+1)\begin{cases} ps_j(n)+\eta(n)h_{j,i(d)}(n)(c_i - ps_j(n)), & \text{se } j \in h_i \\ ps_j(n), & \text{caso contrário} \end{cases}$$ Fórmula 4.6

Onde $ps_j(n+1)$ representa o valor do peso sináptico do neurônio j no tempo imediatamente posterior ao atual. $h_{j,i(d)}(n)$ é o valor da função de vizinhança, $(c_i - ps_j(n))$ é a fórmula para obtenção da taxa de erro e equivale a subtração dos segmentos do vetor de entrada pela vetor de pesos sinápticos do neurônio j, $\eta(n)$ é a taxa de aprendizado (fórmula 4.7), todas no tempo n.

$$\eta(n) = \eta_0 \exp(-\frac{n}{\tau_2})$$ Fórmula 4.7

Onde η_0 é o valor inicial da taxa de aprendizagem, n é um intervalo de tempo discreto, o qual pode ser representado pela iteração corrente em um determinando momento do tempo e τ_2 pode ser dado pelo número total de iterações.

4.4 CARACTERÍSTICAS PARA AGRUPAMENTO E CLASSIFICAÇÃO

O mapa de Kohonen tem a capacidade de realizar tanto o agrupamento quanto a classificação por meio de duas fases: uma de treino e outra de teste (HAYKIN, 2001). A fase de treino é composta por dois estágios: a auto-orga20.136 3u orden.136 e a convergência. Ambos ocorrem no processo adaptativo.

A auto-orga20.136 é responsável pela orde n.136 dos neurônios e ajustes dos pesos sinápticos dos neurônios. A convergência é a responsável por fazer um ajuste fino nos pesos sinápticos dos neurônios, de forma a especializar ainda mais os neurônios para vencerem a competi136 quando os mesmos padrões forem apresentados na entrada. Por meio destas duas etapas é realizado o agrupamento.

Na fase de teste, submete-se à rede trein.da um conjunto de padrões o qual deve ser reconhecid6 e classificado de acordo com o treinamento recebid6. Isto equivale a uma classificação. O mapa treinad6 e testad6 está apto agora para receber outro conjunto de entrada e realizar a classificação. Deve-se notar que o Mapa está especializad6 no contexto da cole136 fornecida como entrada.

Contudo, n36 se pode avaliar de forma precisa os grupos ou classes formadas uma vez que o modelo n36 apresenta uma fronteira bem definida para os grupos formados. Assim, medidas como a *precision* e a *recall*, descritas no capítulo anterior, n36 podem ser utilizadas para os grupos formados no mapa bidimensional.

4.5 AGRUPAMENTO E CLASSIFICAÇÃO DE CRIPTOGRAMAS COM O MAPA DE KOHONEN

A motivação para o us6 do mapa de Kohonen nester trabalh6 foi dada pelo trabalh6 de Souza (2006), no qual descreve um experimento para agrupamento e classifica136 de textos por idioma e assunto, utilizando o mapa de Kohonen. Assim, a mesma rede neural desenvolvida naquele trabalh6 será utilizada na presente disserta136, considerando os ajustes

necessários ao contexto dos criptogramas. O Experimento daquele trabalho será descrito a seguir.

A coleção utiliza naquele experimento foi composta por 210 documentos digitais no formato texto, com sete idiomas/assuntos diferentes, cada um com 30 textos: *alemão*, *francês, inglês, italiano, filosofia* (em português), *direito* (em português) e *medicina natural* (em português). Os textos têm entre 5 e 10 *Kbytes*.

Com o objetivo de verificar de maneira mais clara a formação dos grupos, os textos foram escolhidos considerando, presumivelmente, a baixa interseção entre os mesmos, no caso dos textos escritos em diferentes idiomas e com uma interseção um pouco maior, no caso dos textos escritos em português (classes *filosofia*, *direito* e *medicina natural*). A interseção, no caso destes experimentos, se refere à quantidade de palavras compartilhadas pelos textos das diversas classes. Quanto maior a quantidade de palavras compartilhadas entre as classes, maior a interseção.

A coleção foi dividida em 80% dos documentos para treinamento e os outros 20% para o teste da rede neural.

O resultado é mostrado como um mapa de duas dimensões, onde os documentos são representados com um círculo colorido e a relação geométrica dos círculos indica a similaridade entre os documentos.

Na figura 4.1, pode-se ver o resultado do estágio de treino e na figura 6.2, podemos ver o resultado do estágio de teste.

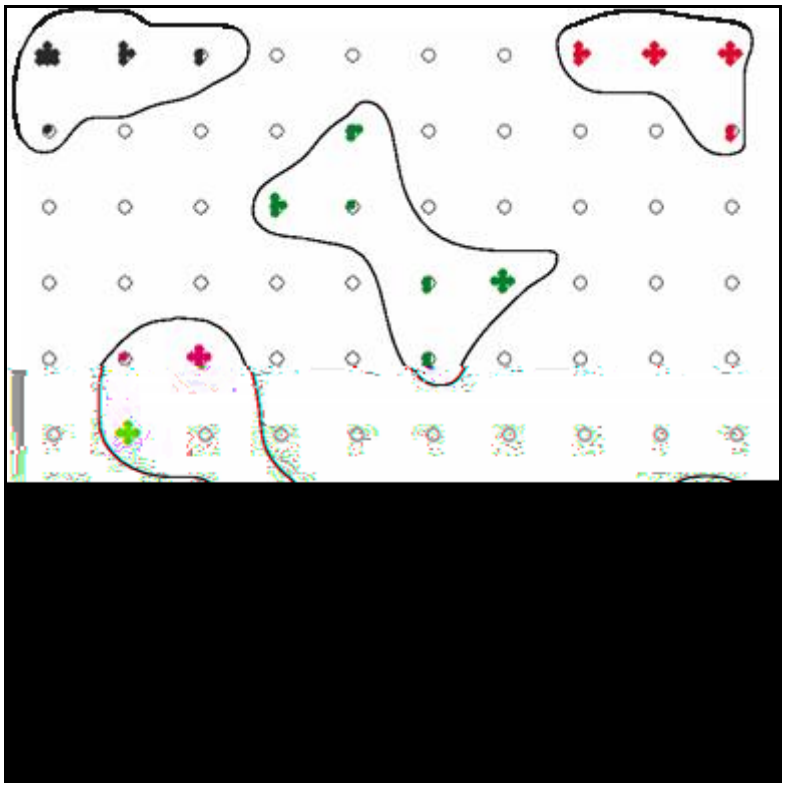


FIG. 4.1: Mapa formado no estágio de treino (SOUZA, 2006)

Por meio da figura 4.1, observa-se que foram formados sete grupos distintos, como esperado, onde cada um desses grupos representa um assunto. As linhas em preto delimitam os grupos formados na fase de treino. As representações nos mapas são feitas por meio de cores, de acordo com a tabela 4.1.

TAB. 4.1: Representação dos idiomas/assuntos no mapa bidimensional

Idioma/Assunto	Cor
Alemão	Azul
Francês	Vermelho
Inglês	Verde
Italiano	Preto
Português (Filosofia)	Amarelo
Português (Direito)	Magenta
Português (Medicina natural)	Cyan

Observando a figura 4.2, vê-se que foram também formados sete grupos distintos, o que também era esperado. Cada um desses grupos representa um assunto. As linhas em preto

delimitam os grupos formados na fase de teste. Note que nesta fase alguns documentos foram mapeados em neurônios diferentes. Contudo, por meio das linhas vermelhas pode-se comparar onde foram mapeados os documentos de assuntos similares na fase de treino. É possível notar que não ocorreram sobreposições, isto é, documentos agrupados em um neurônio na fase de treino, classificados em neurônios que agruparam outros assuntos. Da figura, percebe-se que a precisão permaneceu a mesma.

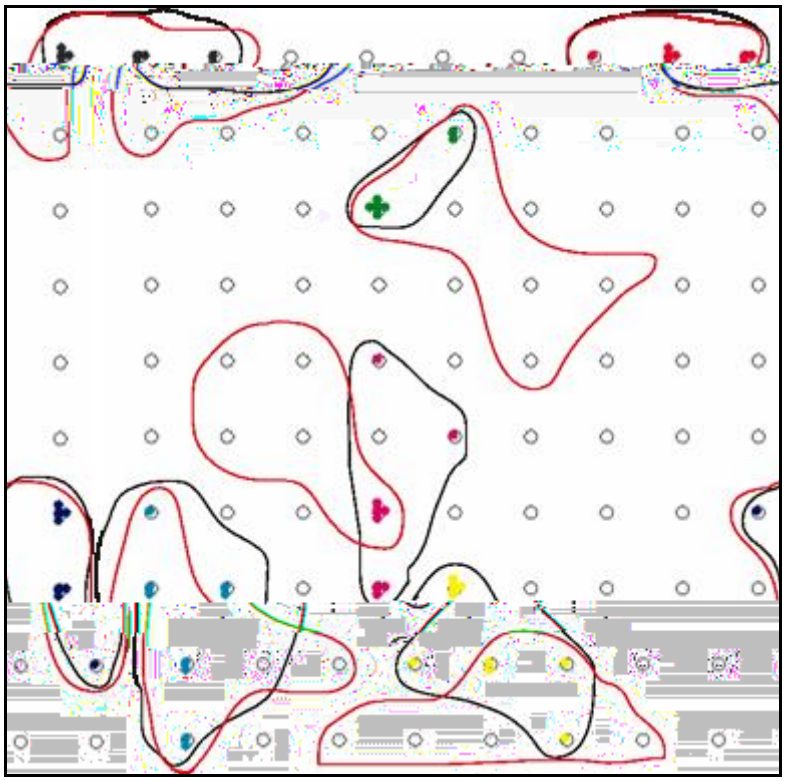


FIG. 4.2: Mapa formado no estágio de teste (SOUZA, 2006)

Naquele trabalho pode ser visto que o agrupamento e a classificação foram realizados com sucesso. Entretanto, este sucesso só pode ser avaliado por meio da inspeção visual ao mapa, não sendo aplicada nenhuma métrica para a avaliação. Nas publicações consultadas, inclusive, escritas e orientadas pelo próprio criador do mapa, não foram encontradas métricas para avaliação do mapa (KOHONEN, 1989), (HONKELA, 1996), (HONKELA^{B}, 1997), (HONKELA^{C}, 1997) e (MERKL, 1997).

Em comparação com o processo de agrupamento por meio dos métodos hierárquicos aglomerativos, a fase inicial de tratamento dos criptogramas (indexação) é idêntica também para o mapa de Kohonen. Contudo, após a indexação, o método hierárquico aglomerativo

efetua os cálculos de similaridades entre todos os criptogramas da coleção gerando uma matriz de similaridades (ou distâncias). Por fim, é realizado o agrupamento por meio de um dos métodos hierárquicos aglomerativos, gerando como saída os grupos com os seus respectivos criptogramas. Já no mapa de Kohonen, após a indexação, o próximo passo é a realização dos três processos descritos anteriomente, para cada criptograma da coleção. Ao final, é gerado um mapa bidimensional mostrando os grupos formados (figura 4.3).

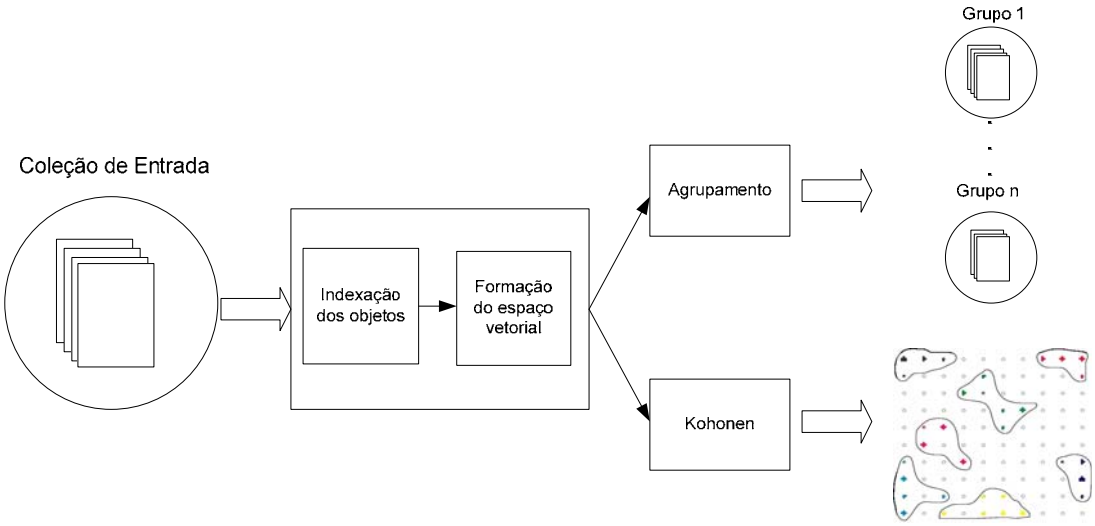


FIG. 4.3: Mapa de Kohonen e Agrupamento hierárquico aglomerativo

Após o agrupamento, o método hierárquico aglomerativo está concluído. Já o mapa de Kohonen pode prosseguir com a classificação.

A execução da rede neural se inicia com um mapa composto por uma camada de neurônios de entrada, representada pelas dimensões dos vetores, que modelam os criptogramas e uma camada de neurônios de saída, onde cada neurônio possui um vetor de pesos sinápticos de dimensão igual à dimensão dos vetores de entrada.

Cada padrão apresentado na entrada (criptograma) é representado por um vetor com as freqüências dos criptotermos. Após os processos competitivo, cooperativo e adaptativo para esse padrão, um novo padrão é submetido ao Mapa. Quando todos os padrões da coleção de entrada já tiverem sido submetidos uma vez ao mapa, está concluída uma época. O número de épocas indica a quantidade de vezes que a coleção de entrada será apresentada ao Mapa. Cada padrão apresentado ao Mapa representa uma iteração.

O tamanho da rede é dado pela indicação linha X coluna. Por exemplo: uma rede 5 X 5, produzirá uma mapa na saída com cinco linhas e cinco colunas e um total de 25 neurônios.

Outro exemplo: uma rede 10 X 10 produzirá um mapa na saída com dez linhas e dez colunas e um total de 100 neurônios.

Para os objetivos deste trabalho, foram realizados alguns experimentos de configuração para verificar quais das medidas representadas pelas fórmulas 4.1 e 4.2 e qual a taxa de aprendizado inicial deveria ser usada. Nas publicações consultadas o valor recomendado para taxa de inicial de aprendizado é de 0,9. Contudo, no trabalho de Souza (2006), a taxa utilizada nos experimentos foi de 0,1. Assim, seria indicado um teste com os criptogramas para se verificar a taxa adequada. A quantidade de épocas foi configura em cinco. Porém, foram feitos experimentos com 10 épocas, mas o resultado não apresentou diferença que justificasse o aumento de tempo necessário para a execução da rede neural. Ressalva-se, mais uma vez, que a avaliação não foi baseada em nenhuma métrica. Desta forma, não pode ser garantido com exatidão o que é o melhor resultado.

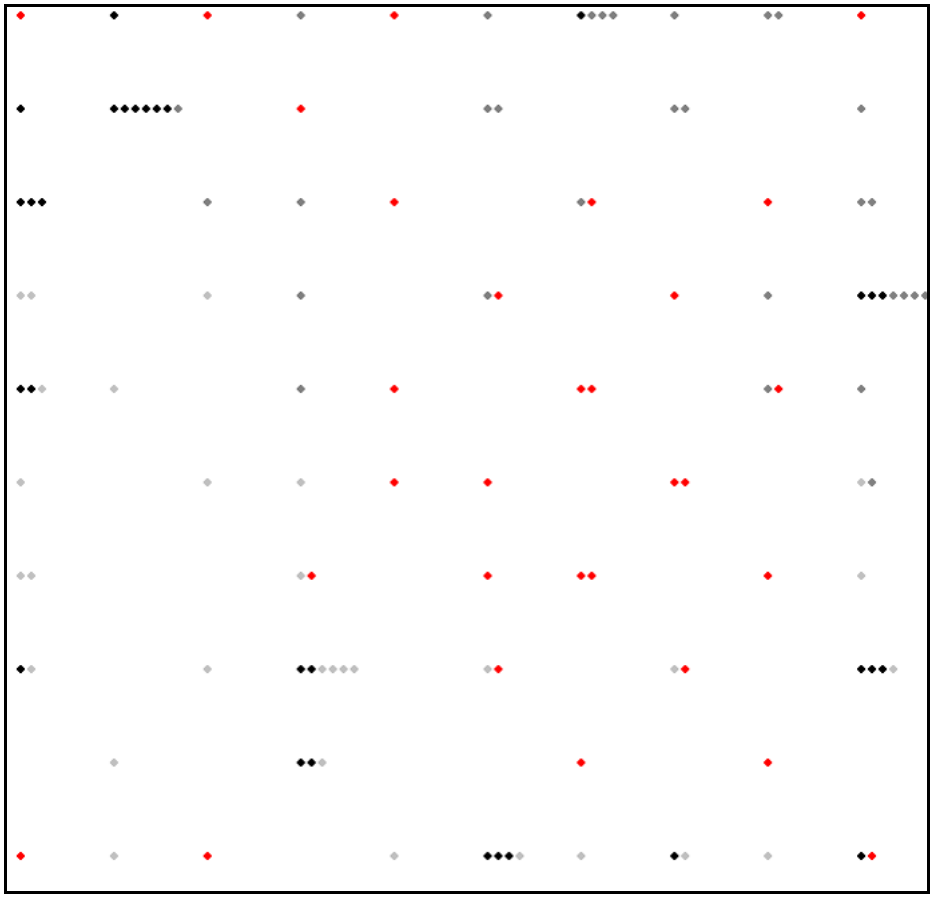


FIG. 4.4: Distância Euclidiana com taxa inicial de aprendizado de 0,9

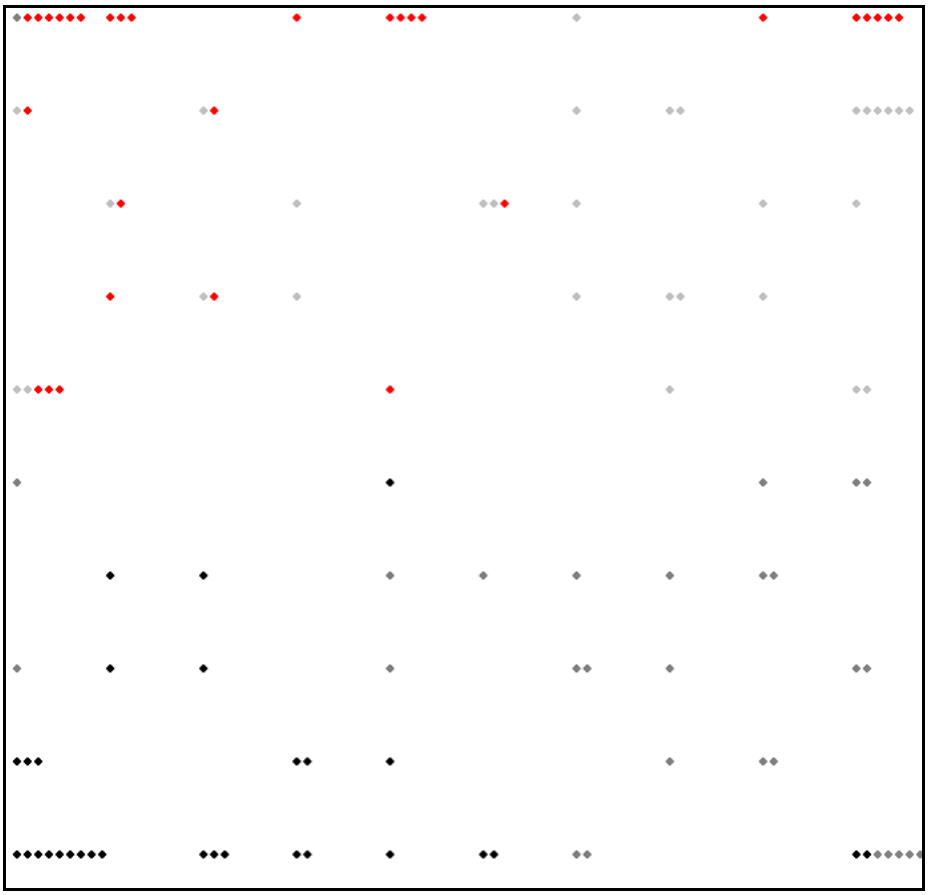


FIG. 4.5: Distância Euclidiana com taxa inicial de aprendizado de 0,1

Para esse experimento de configuração foram utilizados 120 criptogramas, cifrados com quatro chaves diferentes, cada uma cifrando 30 textos. O tamanho dos criptogramas foi de 2048 *bytes*.

As representações nos mapas são feitas por meio de cores, onde os círculos em preto representam os criptogramas cifrados com a chave 1, os círculos em cinza-claro representam os criptogramas cifrados com a chave 2, os círculos em cinza representam os criptogramas cifrados com a chave 3 e os círculos em vermelho representam os criptogramas cifrados com a chave 4.

Na figura 4.4 pode ser visto o resultado do experimento com a distância Euclidiana utilizando a taxa inicial de aprendizado de 0,9 e na figura 4.5, com a taxa inicial de aprendizado de 0,1. Os criptogramas ficaram bastante misturados nos dois mapas. Contudo, visualmente, o mapa que utilizadou a taxa de 0,1 aproximou mais os criptogramas cifrados com a mesma chave, indicando esse valor como mais apropriado.

A figura 4.6 apresenta o resultado do experimento com a medida do ângulo do Co-seno utilizando a taxa inicial de aprendizado de 0,9 e na figura 4.7, com a taxa inicial de aprendizado de 0,1. Os criptogramas obtiveram um melhor agrupamento nos dois mapas, em comparação com os mapas produzidos com a distância Euclidiana. Também visualmente, o mapa que utilizadou a taxa de 0,1 aproximou mais os criptogramas cifrados com a mesma chave, reforçando a indicação desta taxa como a mais apropriada.

Assim, decidiu-se pelo uso da medida do ângulo do Co-seno para o processo competitivo da rede neural artificial e o valor de 0,1 como taxa inicial de aprendizado.

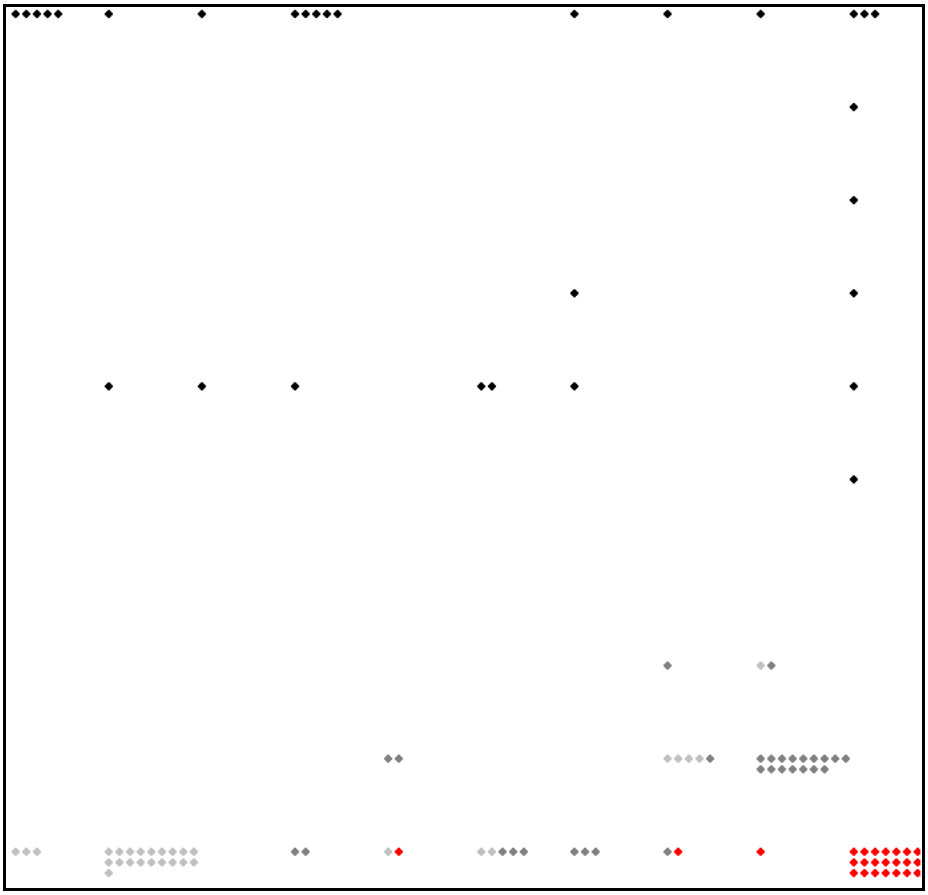


FIG. 4.6: Ângulo do Co-seno com taxa inicial de aprendizado de 0,9

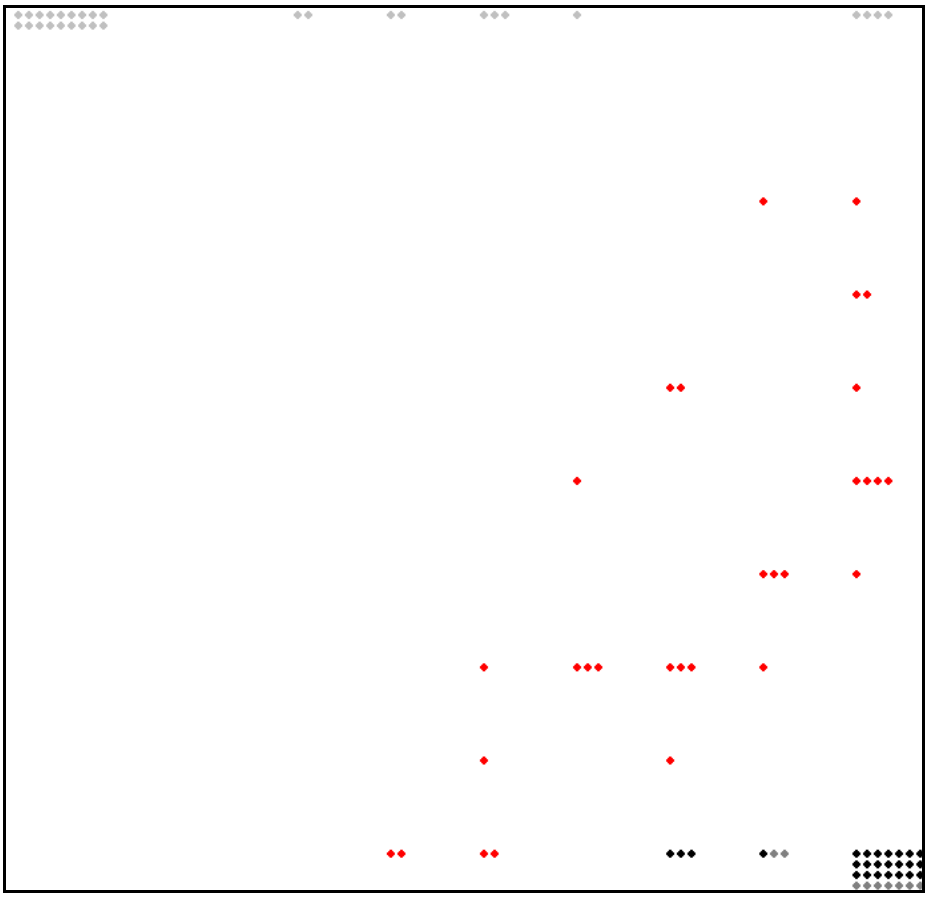


FIG. 4.7: Ângulo do Co-seno com taxa inicial de aprendizado de 0,1

5 FERRAMENTAS

Para a realização dos experimentos deste trabalho, foram desenvolvidas ferramentas e módulos complementares à ferramenta desenvolvida por Carvalho (2006). A linguagem Java foi escolhida para o desenvolvimento.

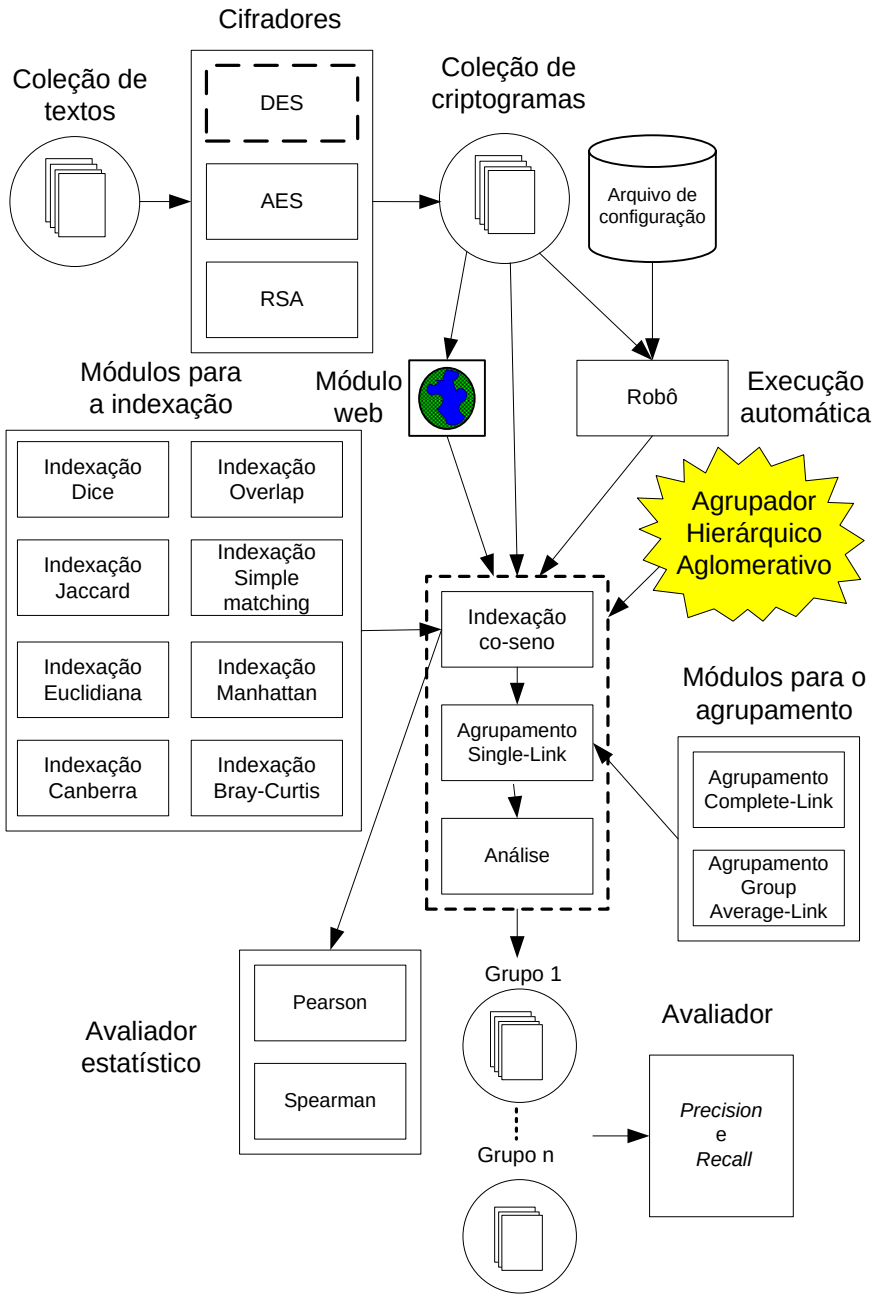


FIG. 5.1: Módulos desenvolvidos para a realização dos experimentos

A figura 5.1 apresenta os detalhes da ferramenta desenvolvida. A parte pontilhada mostra os módulos desenvolvidos em Carvalho (2006), os quais consistem no agrupador hierárquivo aglomerativo propriamente dito e um cifrador para o algoritmo DES. Os demais desenhos mostram os módulos desenvolvidos neste trabalho.

Os novos módulos são descritos abaixo:

- *Avaliador*: um módulo avaliador para calcular automaticamente os valores de precision e recall dos grupos formandos pelo agrupador hierárquivo aglomerativo;
- *Avaliador estatístico*: dois módulos baseados no coeficiente de correlação de Pearson e de Spearman para avaliação das medidas de similaridade e distância;
- *Cifradores*: dois novos módulos cifradores para os algoritmos RSA, chaves de 64, 128, 256, 512, 1024 *bits* (COELHO, 2006), e AES, chaves de 128, 192 e 256 *bits*;
- *Módulos para agrupamento*: dois novos métodos de agrupamento;
- *Módulos para indexação*: quatro novas medidas de similaride e quatro medidas de distância para a indexação;
- *Módulo web*: módulo para submissão via web dos criptogramas para agrupamento (GIRDWOOD, 2006);
- *Robô*: módulo que permite a execução automática de diversas medidas de similaridade e distância e diversos métodos de agrupamento combinados, a partir de um arquivo de configuração;

A figura 5.2 apresenta a ferramenta desenvolvida para os experimentos realizados neste trabalho com a rede neural artificial. Foi desenvolvido um novo indexador para esta ferramenta, embora o indexador utilizado na ferramenta anterior também podesse ser usado. Em testes parciais, o indexador utilizado pela rede neural se mostrou, pelo menos, cinco vezes mais rápido do que o utilizado pelo agrupador hierárquico aglomerativo.

O item tempo de execução é um aspecto importante na aplicação da rede neural artificial. O apêndice 17 apresenta um estudo parcial sobre os tempos necessários para a execução da rede neural, assim como o consumo de memória. O apêndice 18 apresenta as configurações das máquinas utilizadas para o estudo.

6 EXPERIMENTOS, RESULTADOS E AVALIAÇÕES

Dos experimentos realizados em Carvalho (2006) (Figura 6.1), observa-se que dada uma coleção de criptogramas[19] de entrada, utilizando o método de agrupamento hierárquico aglomerativo de ligação simples, são formados grupos precisos[20] e, em alguns casos, com valor máximo de *recall*. Estes grupos formados possuem a característica de que todos os criptogramas membros do grupo foram cifrados com a mesma chave.

Neste trabalho serão considerados os experimentos 1, 3 e 4 de Carvalho, nos Primeiro, segundo e terceiro conjuntos de experimentos, respectivamente.

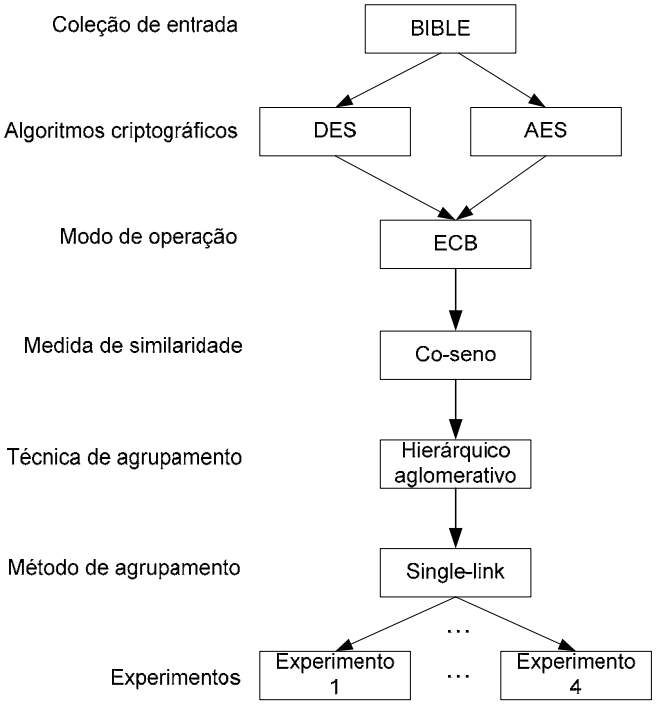


FIG. 6.1: Experimentos realizados em Carvalho (2006)

Neste trabalho, como critério de parada, utilizaremos um valor de similaridade logo acima de zero (valor 0,001), significando que se dois criptogramas possuírem pelo menos um criptotermo em comum eles provavelmente foram cifrados com a mesma chave e deverão ser

[19] Os criptogramas foram cifrados com chaves diferentes por meio dos algoritmos DES e AES, no modo de operação ECB.
[20] Valor máximo para a medida de *precision* (valor 1).

colocados no mesmo grupo. Esta opção visa eliminar uma formação inadequada dos grupos, que pode surgir quando é escolhida uma quantidade de grupos como critério de parada, conforme explicado a seguir.

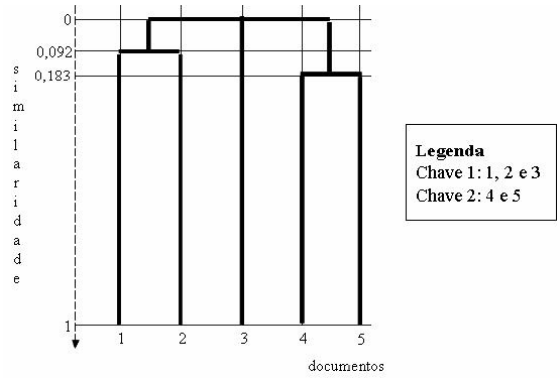


FIG. 6.2: Possível resultado de um processo de agrupamento (CARVALHO, 2006)

Na figura 6.2 pode ser observado que duas chaves distintas foram utilizadas. Contudo, o dendrograma apresenta três grupos, nos quais a similaridade entre criptogramas de grupos diferentes é sempre zero. Desta forma, se o critério de parada for a criação de dois grupos, o algoritmo forçará a união de dois dos três grupos[21], o que provavelmente resultará na mistura de textos cifrados[22] com chaves distintas.

Pode-se observar em Carvalho (2006) que eventualmente ocorrerá que a similaridade entre dois ecriptogramas seja igual a zero, embora tenham sido cifrados com a mesma chave. Assim, esses criptogramas precisam estar no mesmo grupo. Neste caso, a pertinência ao grupo pode ser dada pela co-similaridade com um terceiro criptograma.

A abordagem de se utilizar um valor de similaridade logo acima de zero, como critério de parada, é bastante adequada para as medidas de similaridades uma vez que elas geram valores de similaridades no intervalo [0,1]. Exceção para este intervalo é a medida Simple-matching, cujos valores estão no intervalo [0,∞[. Contudo, as medidas de distância não podem se beneficiar desse critério, já que os seus intervalos de valores estão em [0,∞[. Exceção para este intervalo é a distância Bray-Curtis, cujos valores estão no intervalo [0,1]. Desta forma, para as medidas de distância valores, próximos a zero significam alta similaridade. Assim, é necessário encontrar os maiores valores de dissimilaridades para se

[21] Neste exemplo, a criação de três grupos representa o melhor resultado de agrupamento possível.
[22] Texto cifrado e criptograma são utilizados indistintamente neste trabalho

estabelecer o critério de parada. Esses valores foram encontrados e estão descritos no apêndice 2.

6.2 DESCRIÇÃO DOS EXPERIMENTOS

6.2.1 PRIMEIRO CONJUNTO DE EXPERIMENTOS – INFLUÊNCIA DO TAMANHO DO CRIPTOGRAMA

O objetivo do primeiro conjunto de experimentos é identificar a influência do tamanho do criptograma[23] no resultado do agrupamento. Neste conjunto, são definidos alguns tamanhos de texto, os quais são dependentes do algoritmo criptográfico utilizado. Para cada tamanho, foram extraídos 30 textos da bíblia, os quais foram cifrados usando 50 chaves aleatórias (figura 6.3). Foram utilizados os algoritmos DES, AES e RSA.

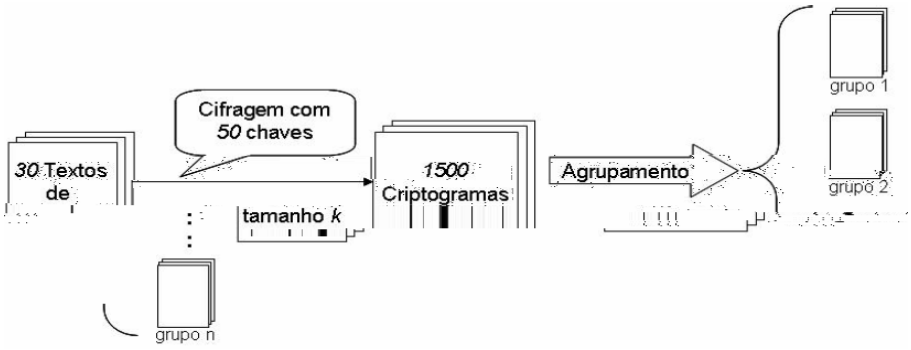


FIG. 6.3: Descrição de um experimento do primeiro conjunto (CARVALHO, 2006)

6.2.1.1 SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO DES

Para o DES, foram definidos 11 tamanhos de textos diferentes, em bytes: 64, 128, 192, 256, 512, 1024, 2048, 4096, 6144, 8192 e 10240, cada tamanho com 30 textos, cifrados com 50 chaves diferentes e aleatórias, totalizando 16.500 criptogramas. Na figura 6.4, pode ser observado uma árvore que descreve o subconjunto de experimentos realizados. Note que cada

[23] Tamanho de criptogramas e tamanho de textos podem ser usados indistintamente neste trabalho, uma vez que tanto um texto em claro como o seu criptograma resultante de um processo de cifragem, para as cifras de blocos aqui consideradas, possuem o mesmo tamanho.

seta leva a uma ramificação. Cada um desses ramos se constitui em uma combinação de itens, a qual define um experimento.

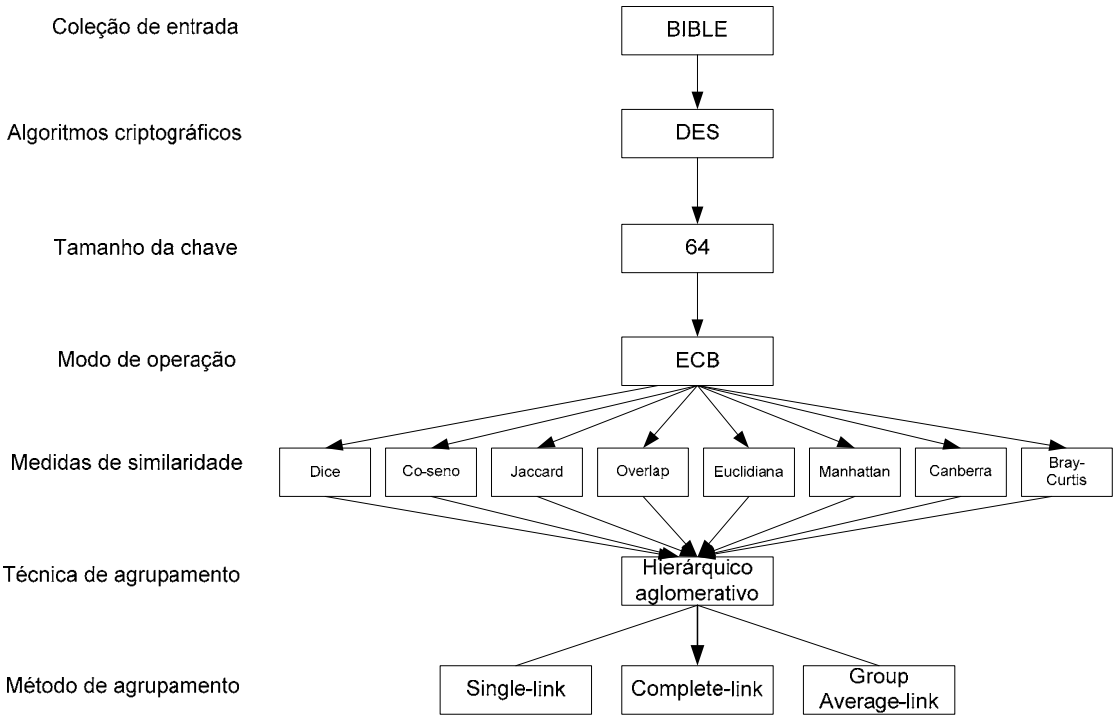


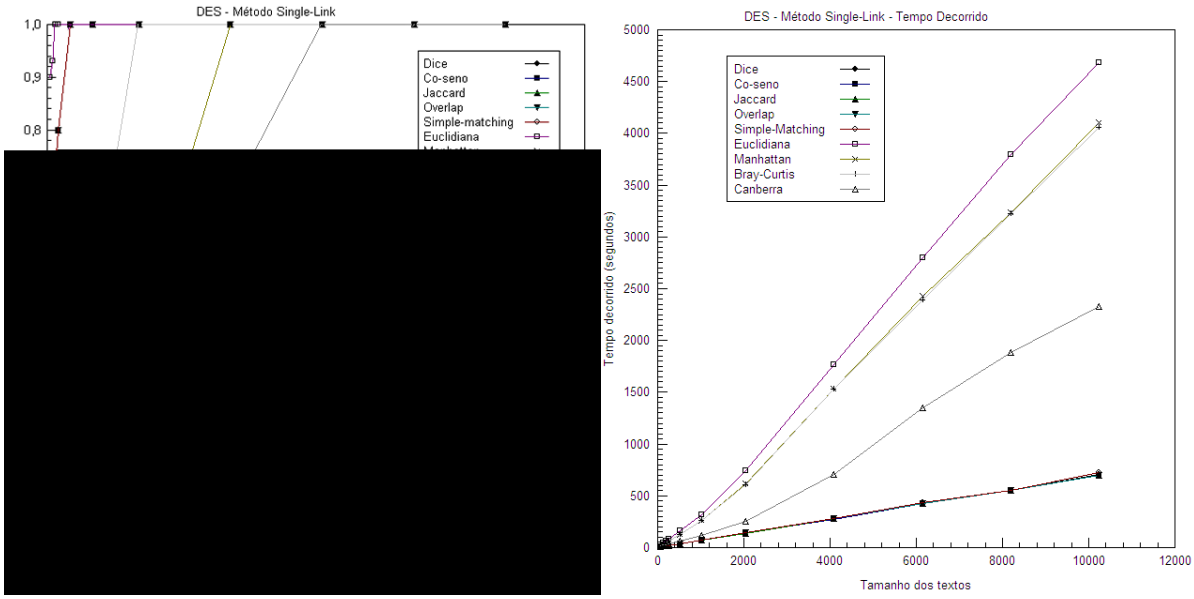
FIG. 6.4: Primeiro conjunto de experimentos: subconjunto para o algoritmo DES

6.2.1.1.1 RESULTADOS E AVALIAÇÕES

O valor da *precision* foi um (1) para todas as medidas de similaridade. Os resultados estão apresentados nos gráficos 6.1, 6.4 e 6.5. Os detalhes podem ser obtidos nas tabelas apresentadas no apêndice 3.

No gráfico 6.1 (a) observa-se que as medidas de similaridade alcançam o valor máximo de *recall* (valor 1) com textos de 512 *bytes* e que mesmo para textos pequenos (64 *bytes*) foram recuperados mais da metade dos textos (indicado pelo valor de *recall* igual a 0,6). Neste caso, a medida de similaridade utilizada não fez diferença. Este resultado é semelhante ao obtido por Carvalho (2006) (tabela 6.1). Assim como naquele trabalho, optou-se por calcular a *recall* de uma determinada chave, tanto para as medidas de similaridade quanto para as de distância, apenas sobre o grupo composto pela maior quantidade de criptogramas (melhor grupo), já que, para alguns tamanhos de textos, muitos grupos podem ser retornados. Essa opção parece adequada, uma vez que provoca um crescimento mais suave do gráfico, se

comparada à medida baseada estritamente na micro-média (Gráfico 6.2 (a) e (b)). Além disso, por meio do Gráfico 6.3, pode-se perceber que o cálculo da micro-média (b) apresenta um crescimento irregular do valor de *recall* para algumas medidas de distância. Porém, quando calculado o melhor grupo, esse resultado se apresenta estável. Desta forma, o grupo com a maior quantidade de criptogramas é considerado o grupo principal, enquanto que os demais grupos podem ser considerados como resíduos do processo de agrupamento.



(a) (b)

GRA. 6.1: *Recall* e tempo decorrido para o algoritmo DES com o método *Single-Link*

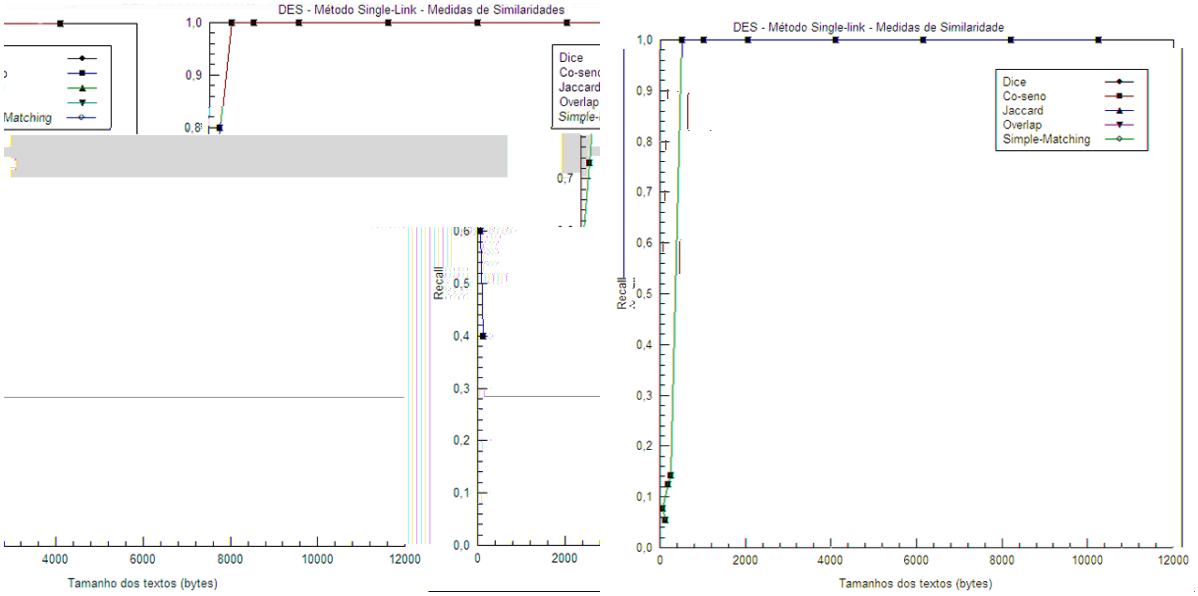
TAB. 6.1: Resultados obtidos em Carvalho (2006) para o algoritmo DES

Tamanho dos Textos (bytes)	Precisão	Abrangência
10240	1	1
8192	1	1
6144	1	1
4096	1	1
2048	1	1
1024	1	1
512	1	1
256	1	0.8
192	1	0.71
128	1	0.4
64	1	0.6

Quanto à *precision*, esta foi calculada considerando todos os grupos formados, tanto para as medidas de similaridade quanto para as de distância.

A partir deste ponto do presente trabalho, todos os valores para *precision* e *recall* serão calculados da maneira descrita nos dois últimos parágrafos.

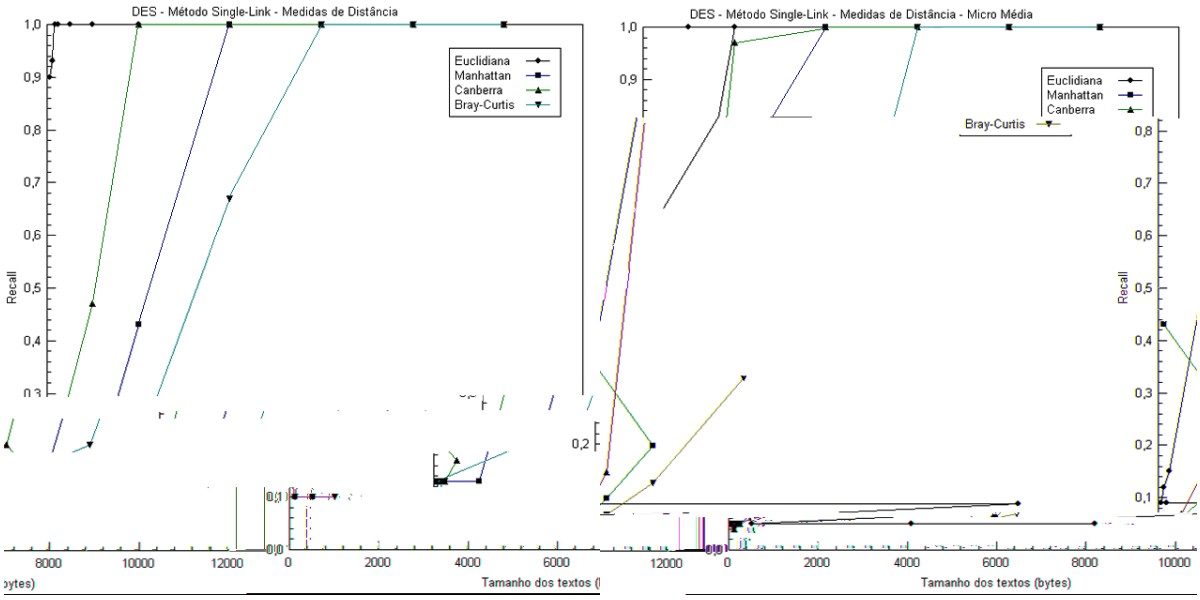
Observando os gráficos 6.4 e 6.5 (a), nota-se que as medidas de similaridades apresentaram um comportamento semelhante também nos métodos *complete-link* e *group average-link*, quando considerados individualmente, excetuando-se para o último caso o coeficiente Jaccard, o qual obteve valor máximo de *recall* a partir de textos com 1024 *bytes*. Contudo, comparando os três métodos de agrupamento, percebe-se que o método *complete-link* teve o pior desempenho, obtendo valor máximo para a *recall* somente a partir de textos com 4096 *bytes*. Já os métodos *single-link* e *group average-link*, tiveram um comportamento semelhante, tendo o *single-link* apresentado um desempenho mais estável.



(a) (b)

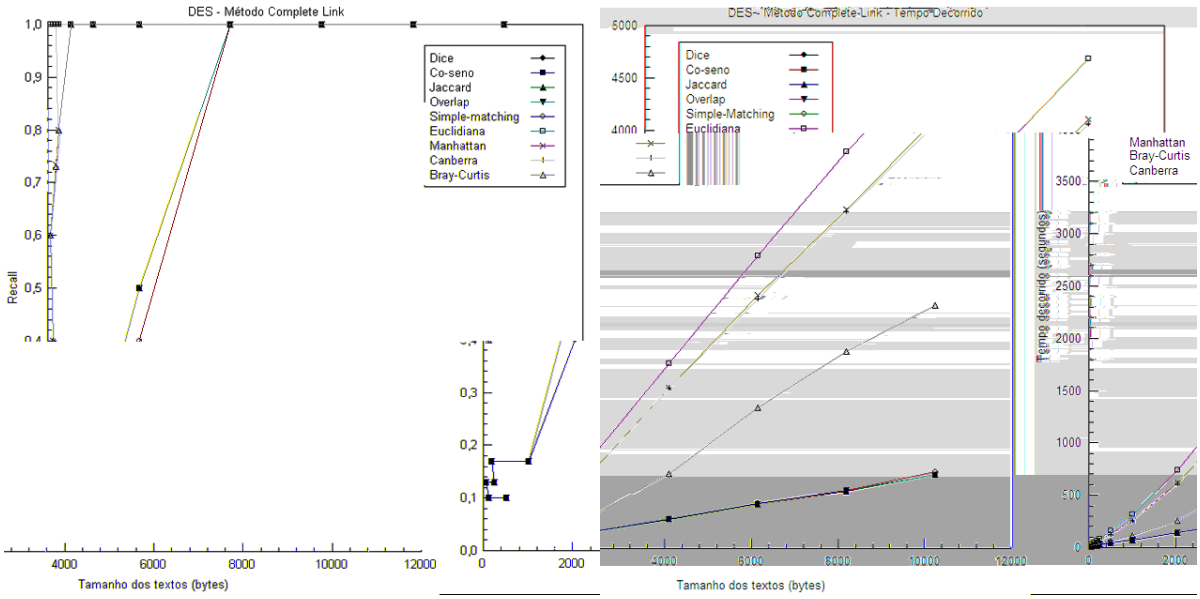
GRA. 6.2: *Recall* para o melhor grupo (a) e micro-média (b) para medidas de similaridade

Os valores relativos ao tempo decorrido nas medidas de similaridades, apresentados no gráfico 6.1, 6.4 e 6.5, (b), tiveram um crescimento suave e compatível com o crescimento do tamanho dos textos e as variações entre as medidas de similaridades foram insignificantes (aproximadamente entre 0 e 10 segundos). No apêndice 18 são apresentadas as configurações dos computadores utilizados para realização dos experimentos.



(a) (b)

GRA. 6.3: *Recall* para o melhor grupo (a) e micro-média (b) para medidas de distância



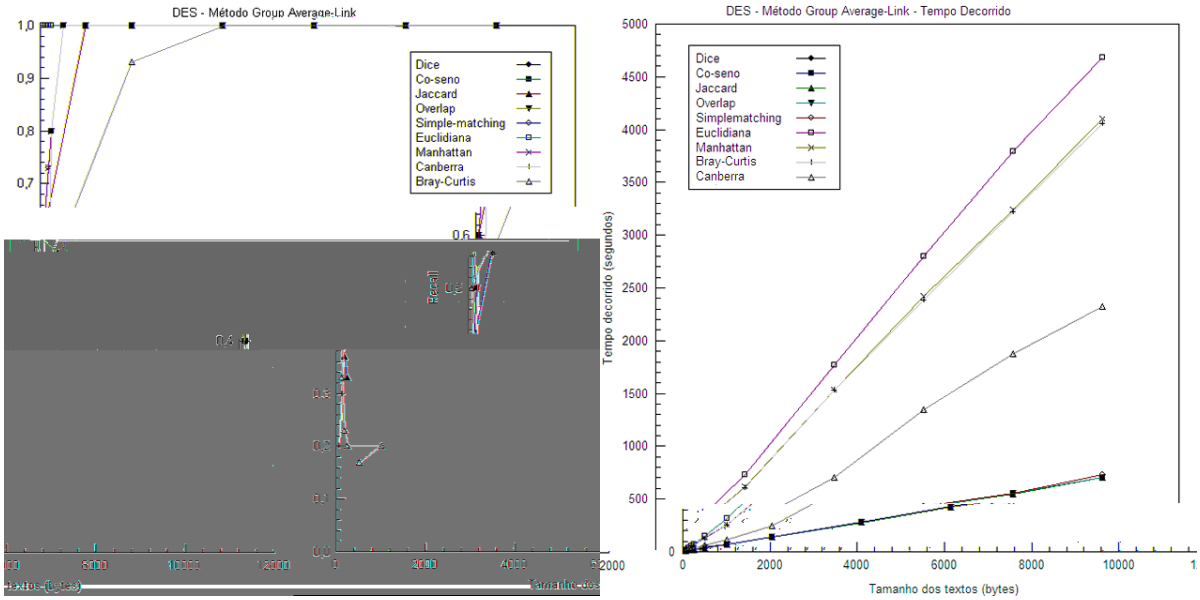
(a) (b)

GRA. 6.4: *Recall* e tempo decorrido para o algoritmo DES com o método *Complete-Link*

Quanto às medidas de distância, como pode ser visto nos gráficos e apêndices, os valores de *precision* neste subconjunto de experimentos para as distâncias Manhattan e Bray-Curtis

foram sempre um. Para a distância Canberra algumas vezes a *precision* não foi igual a um e para a distância Euclidiana, nunca foi um.

Quanto aos valores de *recall*, com o método *single-link*, as distâncias Manhattan e Canberra alcançaram *precision e recall máxima*[24] com textos a partir de 4096 bytes, a distância Bray-Curtis alcançou esses valores com textos a partir de 6144 bytes. A distância Euclidiana foi inadequada, pois embora com *recall* máxima com textos a partir de 192 bytes, mas a sua *precision* não ultrapassou o valor de 0,25; comparando com as medidas de similaridades já relatadas anteriormente, estas alcançaram *precision e recall máxima* neste método, com textos a partir de 512 bytes.



(a) (b)

GRA. 6.5: *Recall* e tempo decorrido para o algoritmo DES com o método *Group Average-Link*

Para o *complete-link*, as distâncias Manhattan e Bray-Curtis foram significativamente melhores, obtendo *recall* máxima com textos de 512 *bytes*; comparando com as medidas de similaridade, estas só chegaram a este valor com textos de 4096 *bytes*. Este resultado sugere que essas medidas de distâncias podem ser equivalentes às medidas de similaridades, quando são considerados os métodos de agrupamento.

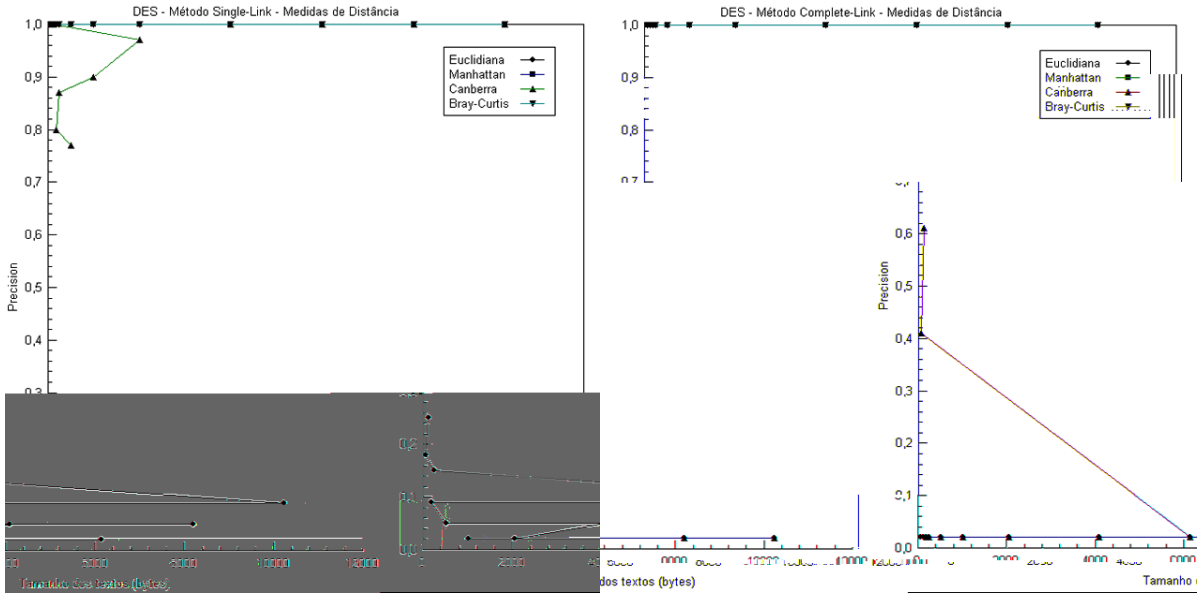
[24] A partir deste ponto no texto, quando for dito "*precision* e *recall máxima*" considera-se os valores em conjunto, ou seja, a medida alcançou para o mesmo tamanho de texto valor máximo para *precision* e para *recall*.

Isto é um comportamento esperado, já que pelo que foi visto até agora o algoritmo *complete-link* funde sempre os grupos com os menores valores de associação[25], o que quer dizer maior similaridade, no caso das medidas de distância e menor similaridade no caso das medidas de similaridades. Assim, o algoritmo *complete-link* privilegia as medidas de distância, em detrimento das medidas de similaridade. No caso do algoritmo *single-link*, este funde os grupos com maiores valores de associação, o que significa que as medidas de similaridade serão privilegiadas em detrimento das medidas de distância. Neste método as distâncias Canberra e Euclidiana também tiveram o valor de *precision* baixo.

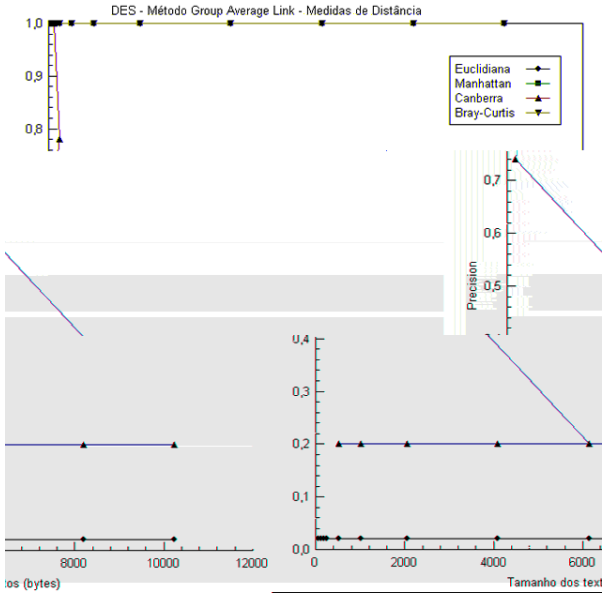
Para o método *group average-link*, a distância Manhattan alcançou *precision e recall máxima* com textos a partir de 512 bytes. A distância Bray-Curtis alcançou esses valores com textos a partir de 4096 bytes. A distância Euclidiana e Canberra tiveram o valor de *precision* baixo.

Destes resultados, conclui-se que o método *single-link* trabalha melhor com as medidas de similaridade, o método *complete-link* trabalha melhor com as medidas de distância Manhattan e Bray-Curtis, e o método *group average-link*, pode lidar bem tanto com as medidas de similaridade, quanto com as medidas de distância Manhattan e Bray-Curtis. As distâncias Euclidiana e Canberra não são adequadas para o uso com os criptogramas gerados pelo DES, pois embora tenham alcançado valores máximos de *recall* em alguns casos, esse resultado é um artefato[26] gerado pela própria incapacidade dessas medidas de criarem grupos bem formados, já que os valores da *precision* foram baixos. Por exemplo, os gráficos 6.6 (b) e (c) mostram que para os métodos *complete-link* e *group average-link*, com a distância Euclidiana, o valor de *precision* ficou em 0,02, para a distância Euclidiana o que significa que apenas um grupo foi formado. A distância Canberra apresentou crescimento irregular, não mostrando qualquer relação entre o tamanho dos textos e os valores de *precision* e *recall*. Os comportamentos das distâncias Euclidiana e Canberra pode ser explicado pelo critério de parada estabelecido para a formação dos grupos (valor de corte), o que forçou o algoritmo de agrupamento a criar grupos com criptogramas cifrados com chaves distintas, diminuindo a precisão.

[25] Similaridade ou distância.
[26] Observação ilusória durante uma medição ou experiência científica e que se deve a imperfeições no método ou na aparelhagem (FERREIRA, 2004).



(a) (b)



(c)

GRA. 6.6: *Precision* para as medidas de distância com os três métodos utilizados

A tomada de tempo, para as medidas de distância, se mostrou bastante alta quando comparada às medidas de similaridade. As distâncias Manhattan e Bray-Curtis consumiram aproximadamente 83% a mais de tempo do que as medidas de similaridades. A distância Canberra obteve o melhor desempenho na tomada de tempo consumindo aproximadamente 70% a mais de tempo do que as medidas de similaridade. A distância Euclidiana obteve o pior desempenho, com aproximadamente 85% de tempo a mais do que as medidas de

similaridades. Comparando a distância Manhattan que foi a melhor melhor distância no agrupamento com a distância Canberra, a qual obteve melhor desempenho no tempo decorrido, a distância Manhattan consumiu 44% tempo a mais do que a distância Canberra.

Este desempenho é parcialmente[27] justificado pela explicação apresentada na seção 3.1.1.2. No apêndice 18 são apresentadas as configurações dos computadores utilizados para realização dos experimentos.

6.2.1.2 EXPERIMENTO COM CRITÉRIO DE PARADA DE 50 GRUPOS COM AS MEDIDAS DE DISTÂNCIA PARA O ALGORITMO DES

Um experimento com as medidas de distância foi necessário sobre os criptogramas do subconjunto de experimentos do algoritmo DES (seção 6.2.1.1). Uma vez que o critério de parada estabelecido (similaridade de corte) forneceu resultado irregular para algumas medidas, o critério de parada foi mudado para a quantidade de grupos, pois sabe-se que a quantidade de grupos ideal a ser formada é de 50 grupos, já te se tem 50 chaves.

Os resultados mostram que com o método *single-link* e as distâncias Manhattan, Bray-Curtis e Canberra, obtiveram *precision e recall máxima* a partir de criptogramas de 4096 *bytes*.

Com o método *group average-link* as distâncias Manhattan e Bray-Curtis, obtiveram *precision e recall máxima* a partir de criptogramas de 512 *bytes*. A distância Canberra obteve *precision e recall máxima* a partir de criptogramas de 4096 *bytes*.

Com o método *complete-link* as distâncias Manhattan e Bray-Curtis, obtiveram *precision e recall máxima* a partir de criptogramas de 512 *bytes*. A distância Canberra obteve *precision e recall máxima* a partir de criptogramas de 6144 *bytes*.

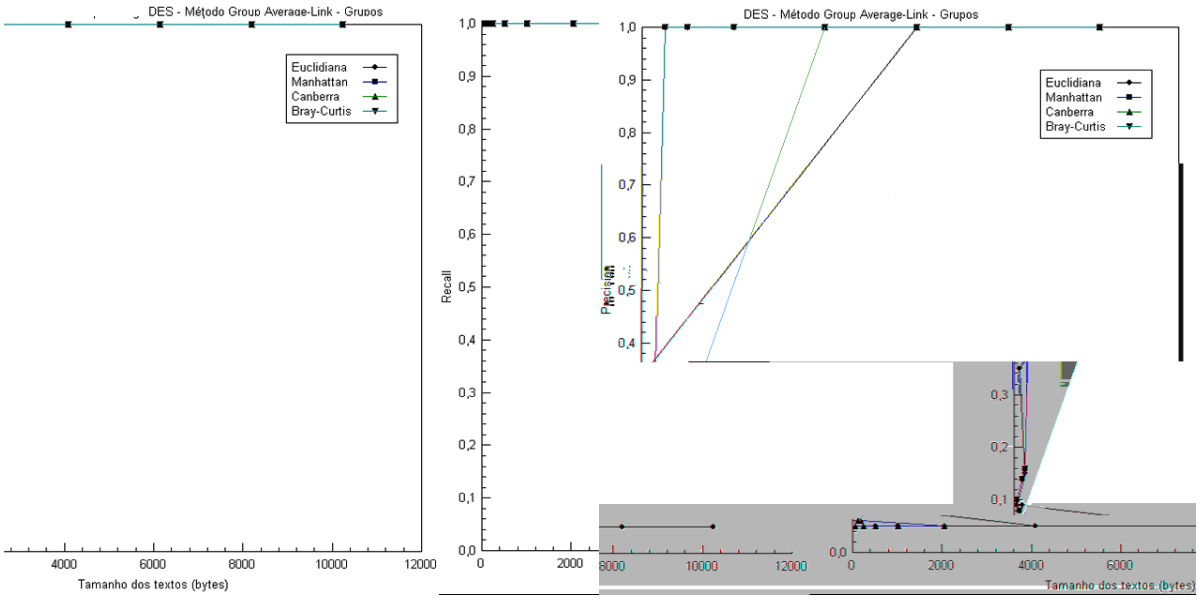
A distância Euclidiana obteve valores máximos somente com criptogramas de 6144 *bytes*, nos métodos *single-link* e *group average-link*. No método *complete-link* a *precision* obteve o valor máximo de 0,14.

O gráfico 6.7 (a) e (b) apresenta os resultados para o método *group average-link*. Os resultados para os outros métodos podem ser vistos no apêndice 3.

Conclui-se que, no caso das medidas de distância, quando se conhece a quantidade de grupos é melhor utilizar esta informação como critério de parada. Deve-se avaliar

[27] O valor da tomada de tempo poder ser afetado por diversas outras características, como os computadores utilizados e o estilo de programação.

cuidadosamente o uso da distância Euclidiana, já que em apenas dois caso a mesma alcançam valores adequados.



(a) (b)

GRA. 6.7: *Precision* e *recall* para o método *Complete-Link* com 50 grupos

6.2.1.3 SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO AES

Para o AES, foram definidos nove tamanhos de textos diferentes, em *bytes*: 1024, 1536, 2048, 2560, 3072, 4096, 6144, 8192 e 10240, cada tamanho com 30 textos, cifrados com 50 chaves diferentes e aleatórias, utilizados três tamanhos diferentes de chaves (150 chaves aleatórias), totalizando 40.500 criptogramas. Na figura 6.5, pode ser observado uma árvore que descreve os experimentos realizados. Note que cada seta leva a uma ramificação. Cada um desses ramos se constitui em uma combinação de itens, a qual define um experimento.

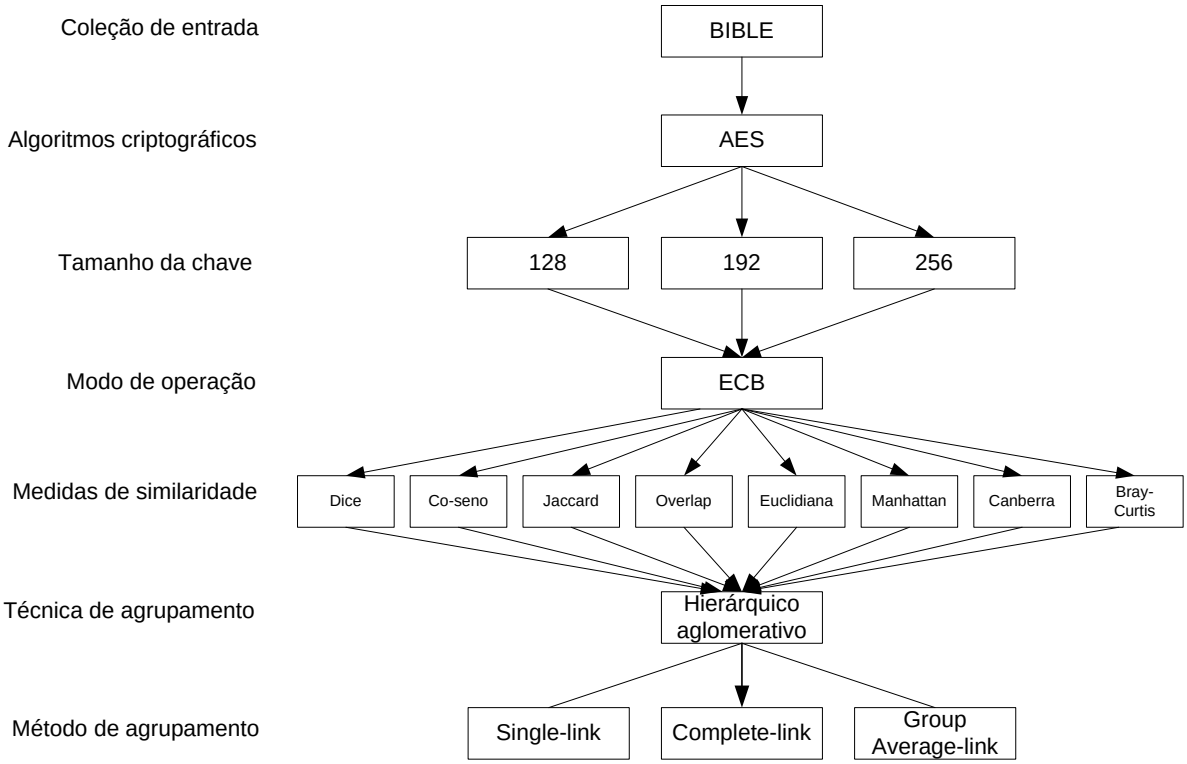


FIG. 6.5: Primeiro conjunto de experimentos: subconjunto para o algoritmo AES

6.2.1.3.1 RESULTADOS E AVALIAÇÕES

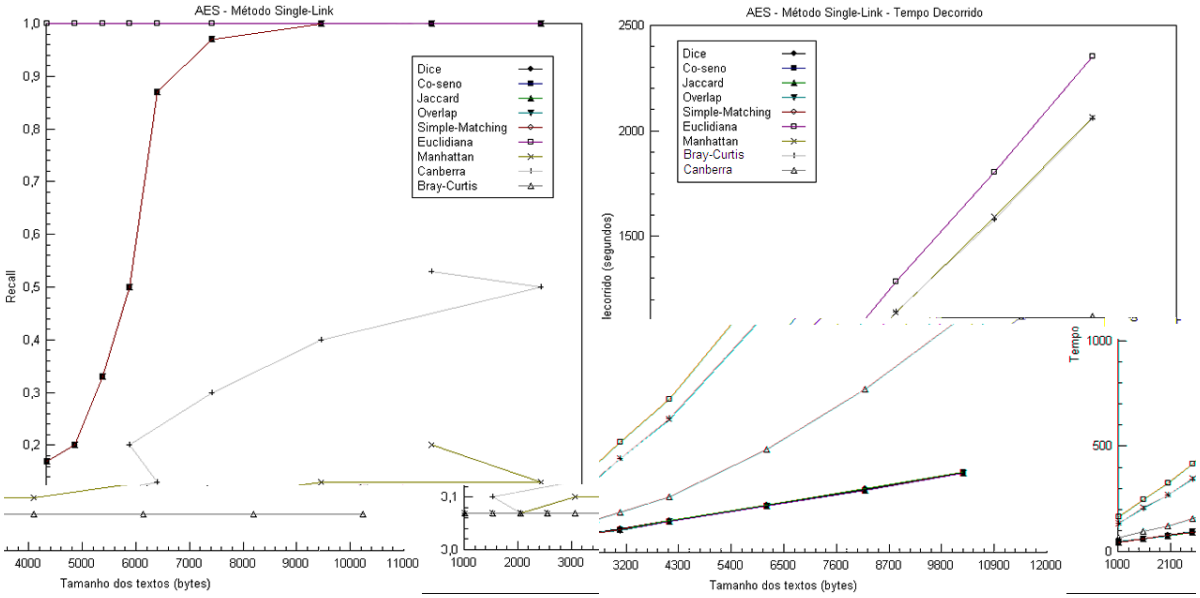
A chave determina um alfabeto e o tamanho da chave indica o espaço de criptotermos possíveis de ocorrerem em um criptograma. Desta forma, no caso do algoritmo AES, teremos um espaço de termos de 2^{128} para a chave de 128 bits, 2^{192} para a chave de 192 bits e 2^{256} para a chave de 256 bits. Disto, pode-se inferir que o tamanho dos criptogramas necessários para realizar o agrupamento com máximas *precision* e *recall* devem ser maiores que os cifrados com o algoritmo DES, o qual possui um espaço de termos de 2^{64}.

De fato, como pode ser observado nos gráficos 6.8, 6.9 e 6.10 e no apêndice 4, com chaves de 128 bits, para as medidas de similaridade, nos métodos *single-link* e *group average-link*, só foi possível alcançar *precision e recall máxima* com textos a partir de 6144 bytes. Com o método *complete-link*, nenhuma das medidas de similaridade alcançou estes valores. Os resultados estão compatíveis com os obtidos em Carvalho (tabela 6.2).

TAB. 6.2: Resultados obtidos em Carvalho (2006) para o algoritmo AES

Tamanho dos Textos (bytes)	Precisão	Abrangência
10240	1	1
8192	1	1
6144	1	1
4096	1	0.97
3072	1	0.87
2560	1	0.5
2048	1	0.33
1536	1	0.2
1024	1	0.16

Já as medidas de distância, nos métodos *single-link* e *group average-link*, só foi possível alcançar *precision e recall máxima* neste último método e com a distância Manhattan, para textos de 10240 bytes. Com o método *complete-link*, a distância Manhattan alcançou estes valores para textos a partir de 6144 bytes.

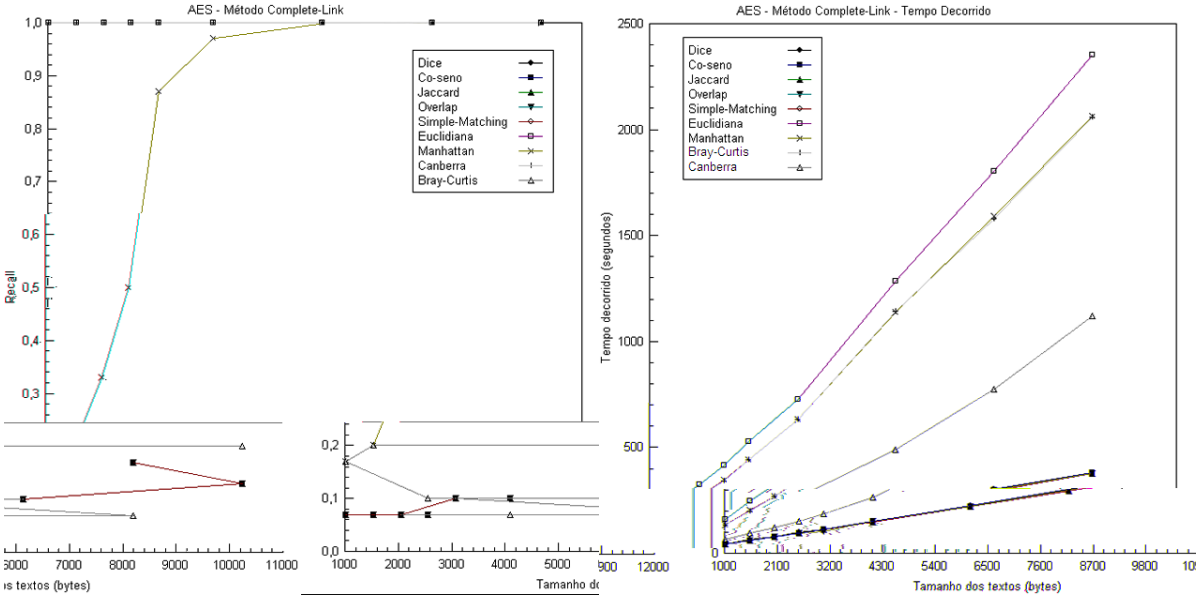


(a) (b)

GRA. 6.8: *Recall* e tempo decorrido para o algoritmo AES com, com chaves de 128 *bits*, o método *Single-Link*

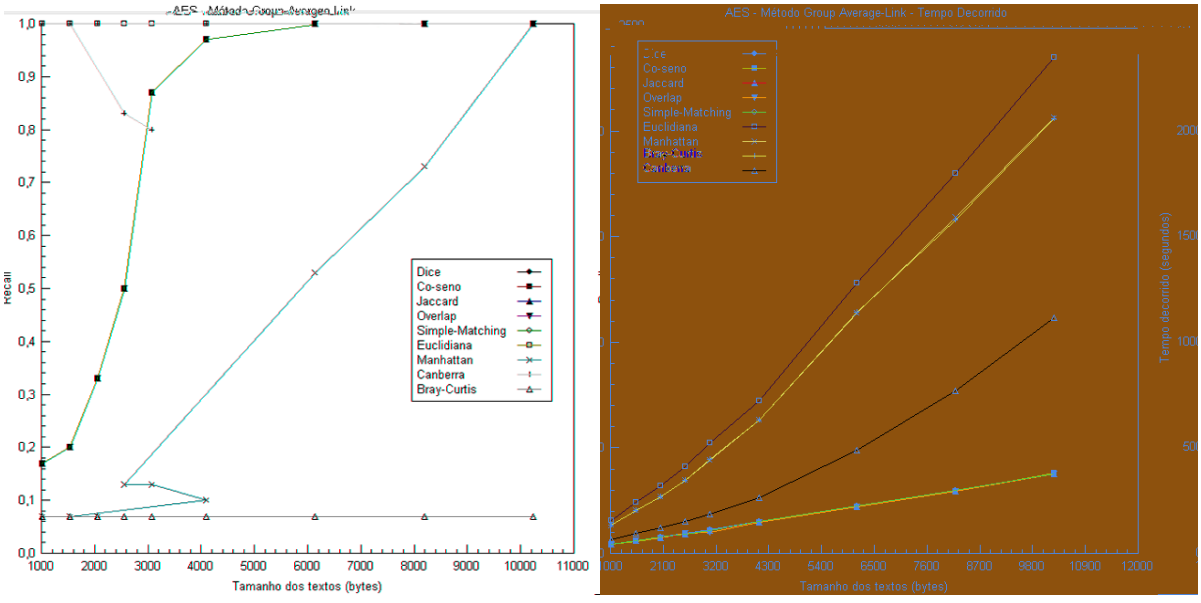
Os valores relativos ao tempo decorrido nas medidas de similaridades e distância, apresentados no gráfico 6.8, 6.9 e 6.10, (b), mantiveram comportamento semelhante ao

apresentado para o algoritmo DES e compatível com o tamanho dos textos. A exceção feita ao valor dos tempos, os quais foram menores.



(a) (b)

GRA. 6.9: Recall e tempo decorrido para o algoritmo AES, com chaves de 128 *bits*, com o método *Complete-Link*



(a) (b)

GRA. 6.10: *Recall* e tempo decorrido para o algoritmo AES, com chaves de 128 *bits*, com o método *Group Average-Link*

Com chaves de 192 e 256 não foi possível alcançar *precision e recall máxima* em nenhum dos métodos com nenhuma das medidas de similaridade e distâncias. Os valores de *precision* e *recall*, quando considerados isoladamente, permaneceram como semelhantes aos valores obtidos nos experimentos anteriores.

Como esperado, o valor de *recall* vai melhorando à medida que os tamanhos de textos aumentam. Apenas um comportamento fora do padrão foi percebido. No caso das chaves de 256 *bits*, com as medidas de similaridade, nos métodos s*ingle-link* e g*roup average-link*, para textos de 8192 e 10240 *bytes* obteve valores de *recall* melhores (0,17 e 0,25, respectivamente) do que para as chaves de 192 *bits* com os mesmos tamanhos de textos (0,13 e 0,17, respectivamente). Neste caso, o esperado que os textos cifrados com chaves de 192 *bits* obtivessem resultados melhores, dado ao explicado no início desta seção.

Destes resultados, conclui-se que o método *single-link* trabalha melhor com as medidas de similaridade, o método *complete-link* trabalha melhor com as medidas de distância Manhattan e Bray-Curtis, e o método *group average-link*, pode lidar bem tanto com as medidas de similaridade, quanto com a distância Manhattan. As distâncias Euclidiana e Canberra não foram adequadas para o uso com os criptogramas gerados pelo AES, já que os valores da *precision* foram baixos. Os comportamentos das distâncias Euclidiana e Canberra pode ser explicado pelo critério de parada estabelecido para a formação dos grupos (valor de corte), o que forçou o algoritmo de agrupamento a criar grupos com criptogramas cifrados com chaves distintas, diminuindo a precisão.

Os resultados completos podem ser vistos nos apêndices 5 e 6.

6.2.1.4 SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO RSA

Para o RSA, foram definidos 11 tamanhos de textos diferentes, em bytes: 64, 128, 192, 256, 512, 1024, 2048, 4096, 6144, 8192 e 10240, cada tamanho com 30 textos, cifrados com 50 chaves diferentes e aleatórias, utilizados cinco tamanhos diferentes de chaves (250 chaves aleatórias), totalizando 82.500 criptogramas. Na figura 6.6, pode ser observada uma árvore que descreve os experimentos realizados. Note que cada seta leva a uma ramificação. Cada um desses ramos se constitui em uma combinação de itens, a qual define um experimento.

Os cinco tamanhos de chaves utilizados foram 64, 128, 256, 512 e 1024 *bits*. Embora os tamanhos de chaves utilizados com o algoritmo RSA estejam normalmente entre 512 e 2048 bits (ODLYZKO, 1995) e (RSA, 2006), este experimento utiliza tamanhos de chaves

menores do que os recomendados, para testar se as técnicas utilizadas neste trabalho são válidas, em algum momento, para o RSA, pois como visto nos experimentos com o AES, chaves a partir de 192 bits alcançaram valor máximo de 0,23 para a *recall* (apêndices 5 e 6).

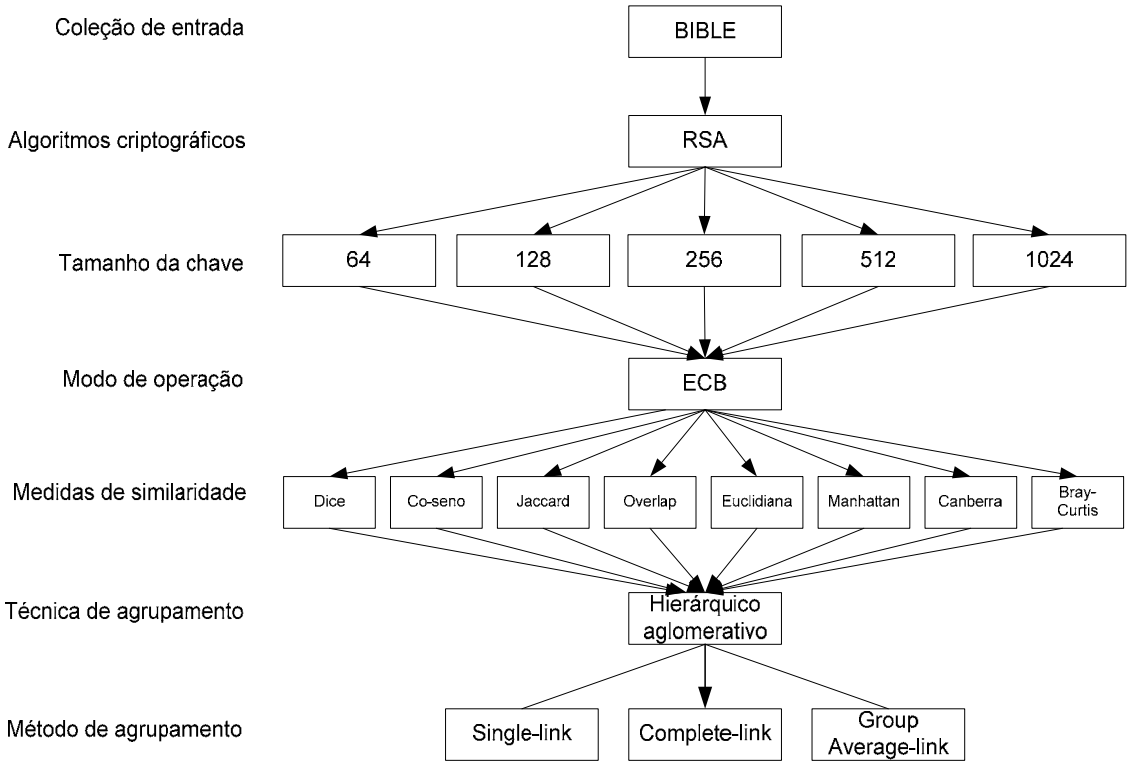
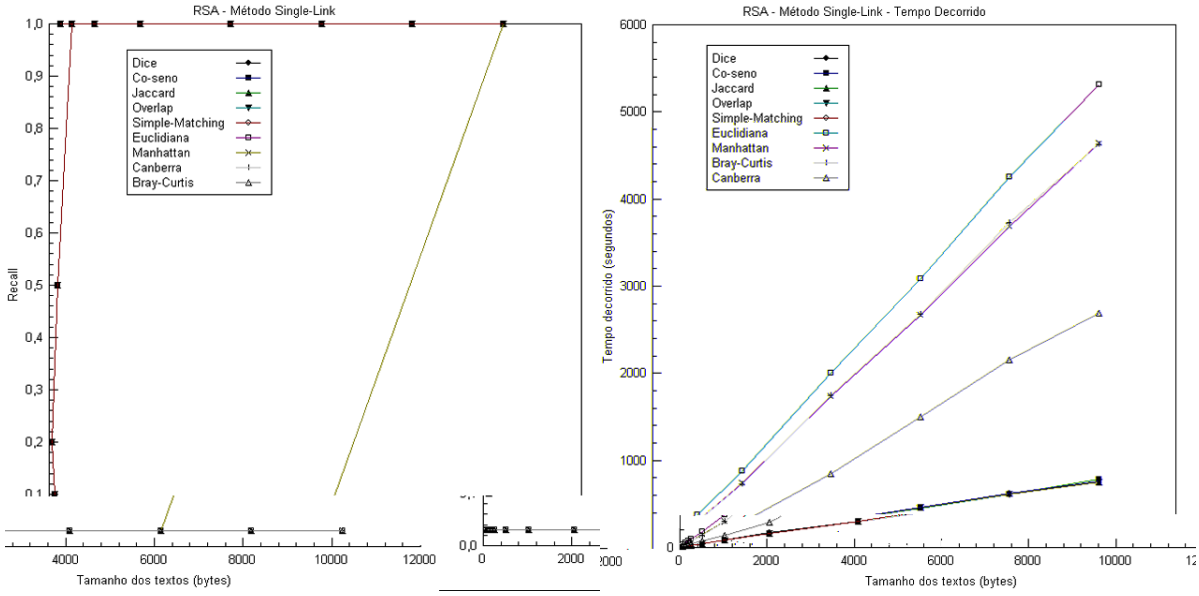


FIG. 6.6: Primeiro conjunto de experimentos: subconjunto para o algoritmo RSA

6.2.1.4.1 RESULTADOS E AVALIAÇÕES

A observação feita para o algoritmo AES sobre o tamanho dos criptogramas necessários para realizar o agrupamento com máximas *precision* e *recall*, também são válidas para o RSA. Neste caso, teremos espaços de termos que vão de 2^{64} até 2^{1024}.



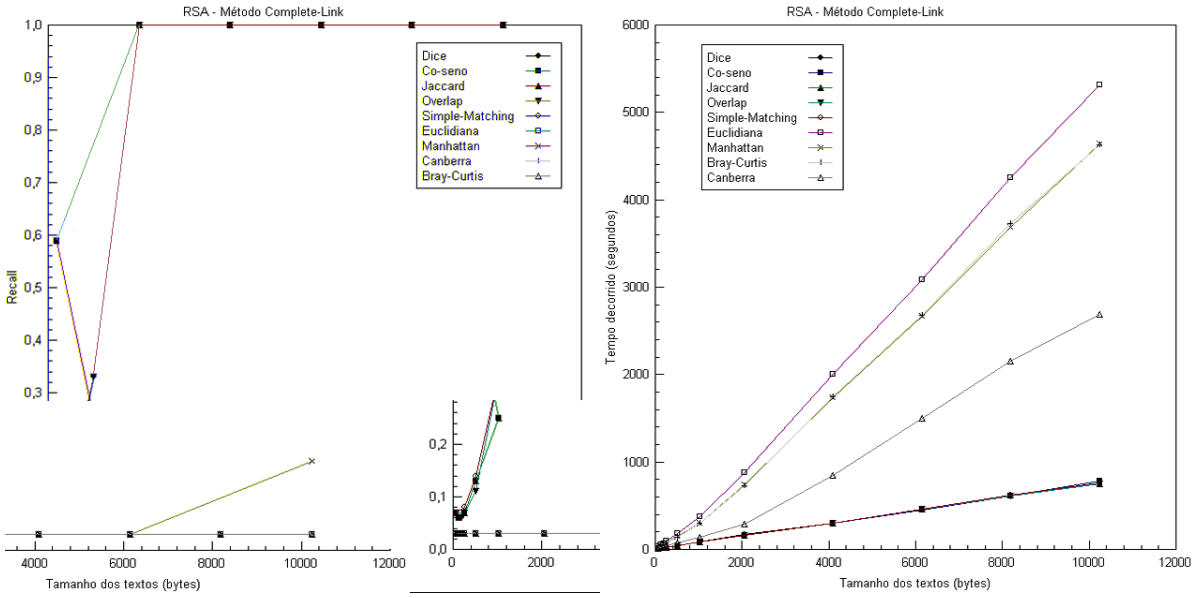
(a) (b)

GRA. 6.11: *Recall* e tempo decorrido para o algoritmo RSA com, com chaves de 64 *bits*, com o método *Single-Link*

De uma maneira geral, os métodos de agrupamento e as medidas de similiraridade e distância produziram resultados semelhantes aos dos experimentos anteriores. Nos gráficos 6.11 a 6.13, (a) e no apêndice 7, estão os resultados para chaves de 64 bits. Para estas chaves , observa-se que as medidas de similaridade, nos métodos *single-link* e *group average-link*, alcançaram *precision e recall máxima* com textos a partir de 256 *bytes*. Com o método *complete-link*, estas medidas alcançaram estes valores com textos a partir de 2048 *bytes*. Os valores de *precision* foram um em todos os casos.

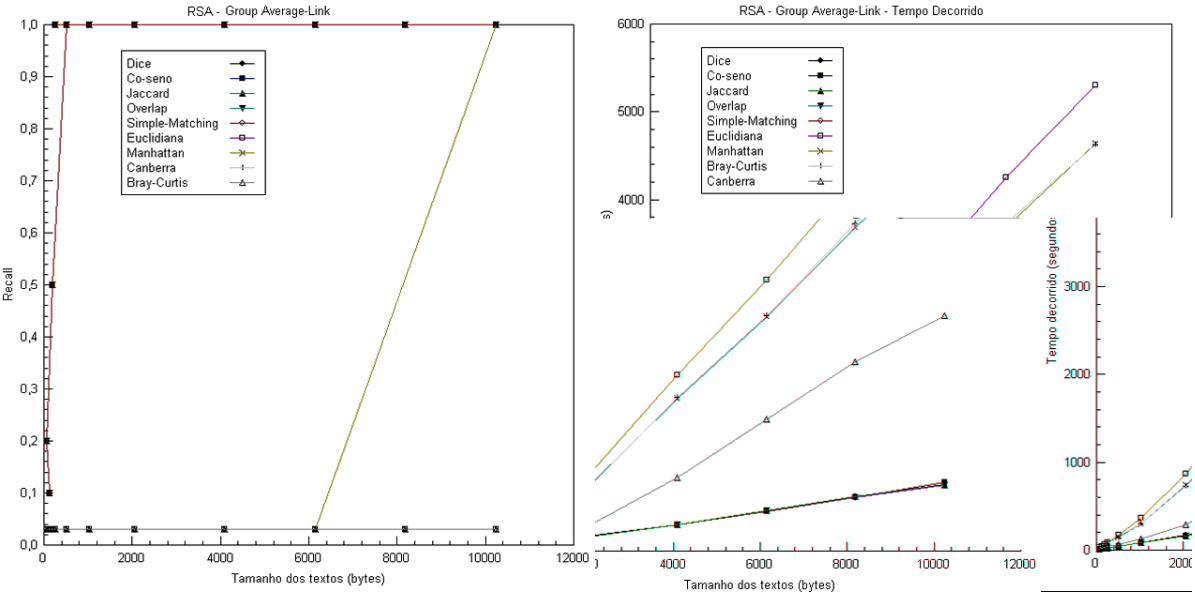
Já as medidas de distância, tiveram resultados pouco expressivos, do ponto de visgta da *recall*. O maior valor de *recall*, com *precision* um, foi 0,03, o que significa que foram formados 1500 grupos, cada grupo com um criptograma. Contudo, a precision foi um em quase todos os experimentos, só ficando abaixo de um em dois experimentos com a distância Manhattan e em dois experimentos com a distância Euclidiana.

Os valores relativos ao tempo decorrido nas medidas de similaridades e distância, apresentados nos gráficos 6.11 a 6.13, (b), foram semelhantes aos experimentos anteriores e apresentaram crescimento compatível com o tamanho dos textos.



(a) (b)

GRA. 6.12: *Recall* e tempo decorrido para o algoritmo RSA com, com chaves de 64 *bits*, com o método *Complete-Link*

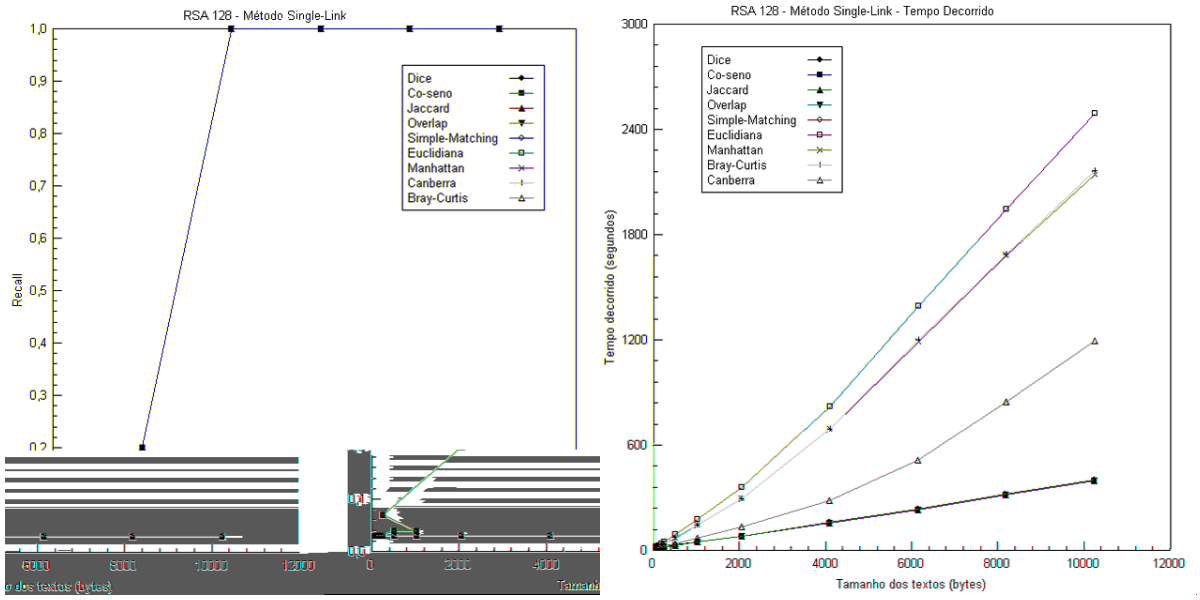


(a) (b)

GRA. 6.13: *Recall* e tempo decorrido para o algoritmo RSA com, com chaves de 64 *bits*, com o método *Group Average-Link*

O valor da *precision* foi um para todas as medidas. Nos gráficos 6.14 a 6.16, (a) e no apêndice 8, podem ser vistos os resultados para as chaves de 128 bits, nos quais as medidas de similaridade, nos métodos *single-link* e *group average-link*, alcançaram *precision e recall*

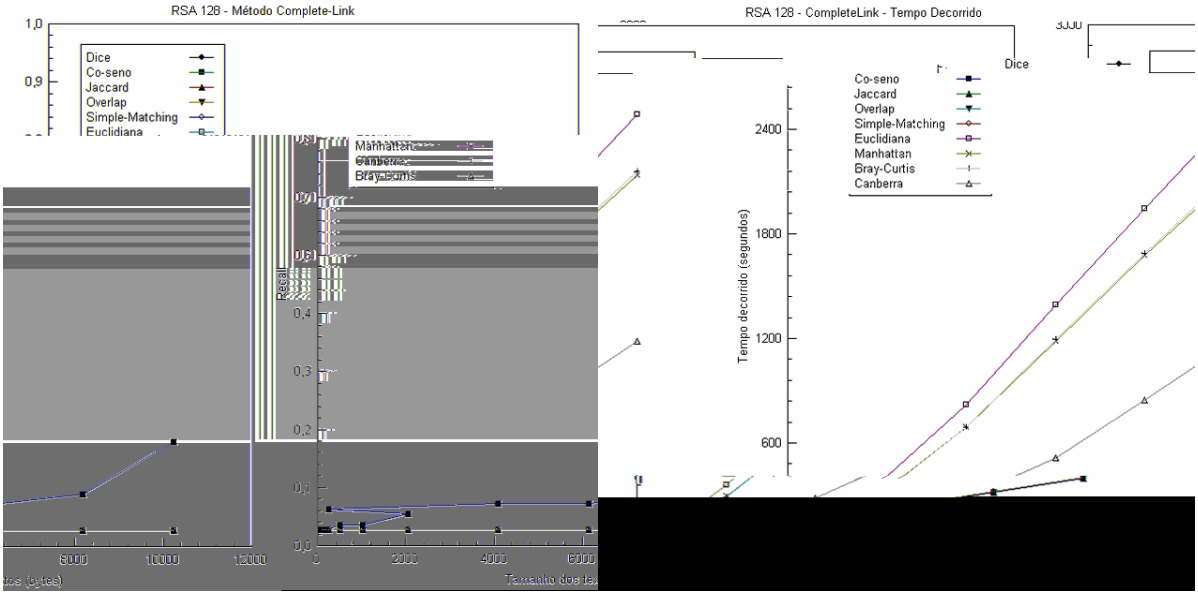
máxima com textos a partir de 4096 *bytes*. Com o método *complete-link*, estas medidas alcançaram o valor máximo de 0,17 para a *recall*. As medidas de distância tiveram como maior valor de *recall* 0,03, o que significa que foram formados 1500 grupos, cada grupo com um criptograma.



(a) (b)

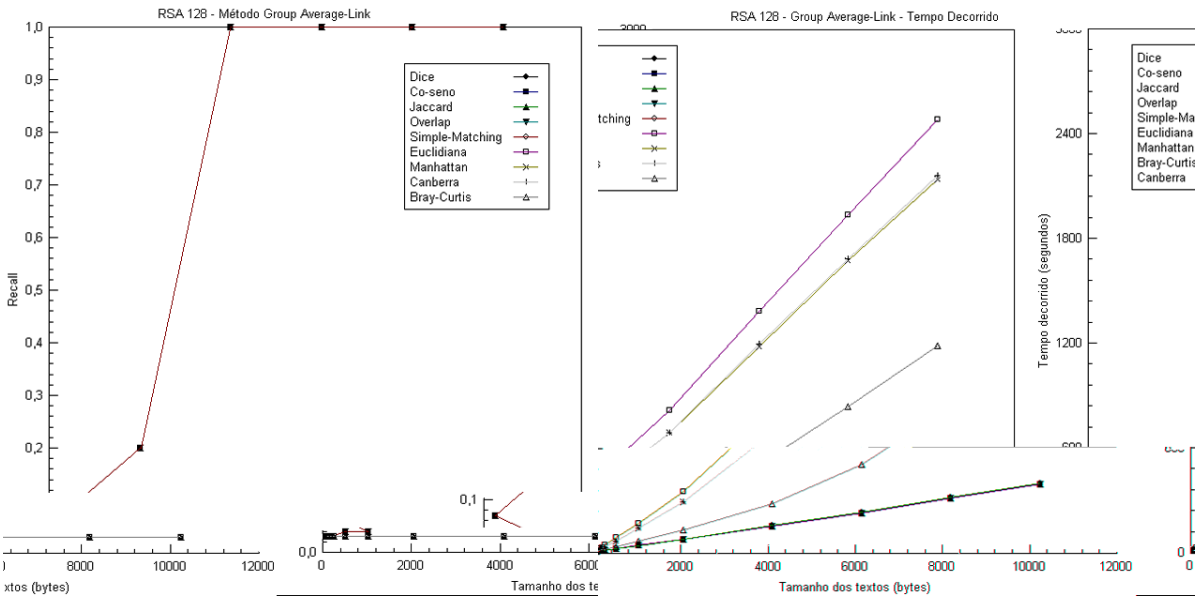
GRA. 6.14: *Recall* e tempo decorrido para o algoritmo RSA com, com chaves de 128 *bits*, com o método *Single-Link*

Os valores relativos ao tempo decorrido nas medidas de similaridades e distância, apresentados nos gráficos 6.14 a 6.16, (b), foram semelhantes aos experimentos anteriores e apresentaram crescimento compatível com o tamanho dos textos.



(a) (b)

GRA. 6.15: *Recall* e tempo decorrido para o algoritmo RSA com, com chaves de 128 *bits*, com o método *Complete-Link*



(a) (b)

GRA. 6.16: *Recall* e tempo decorrido para o algoritmo RSA com, com chaves de 128 *bits*, com o método *Group Average-Link*

Com chaves de 256, 512 e 1024 *bits* o valor de *precision* foi um para todos os experimentos. Mas não foi possível alcançar *precision e recall máxima* em nenhum dos métodos com nenhuma das medidas de similaridade e distâncias. O máximo valor de *recall*

alcançado foi 0,08 para chaves de 256 *bits*, 0,06 para chaves de 512 *bits* e 0,07 para chaves de 1024 *bits*, com criptogramas de até 10240 *bytes*. Foi realizado um novo experimento com criptogramas entre 12240 e 20160 *bytes*, sendo alcançado o valor de *precision* um em todas as medidas de similaridade em todos os métodos e com a distância Euclidiana e Canberra, nos métodos *single-link* e *group average-link* e o valor máximo de *recall* de 0,67 (ver seção 6.2.6, para maiores detalhes). Os resultados completos podem ser vistos nos apêndices 9, 10, 11 e 16.

Para este experimento, conclui-se que os métodos *single-link* e *group average-link* trabalham melhor com as medidas de similaridade, o método *complete-link* não obteve bons resultados, chegando ao valor máximo de 0,20 para *recall*. Todas as medidas de distâncias obtiveram recall 0,03 na maioria dos experimentos deste subconjunto, indicando a formação de grupos com apenas um criptograma em cada grupo.

Comparando os resultados obtidos com o algoritmo RSA, chave de 64 bits e 128 bits, com o algoritmo DES, chave 64 bits, e o AES, chave de 128 bits, respectivamente, pode se observar que para as medidas de similaridade os valores relativos a recall foram melhores, isto é, o RSA obteve valores máximos de precision e recall com criptogramas menores do que o DES e o AES. Já com as medidas de distância, os resultados pioraram ainda mais para o RSA, pois em alguns casos com o DES e o AES foram alcançados *precision e recall máxima*. Já para o RSA isto não aconteceu em nenhum momento.

Com o espaço de termos de 2^{256} (chaves de 256 *bits*), o AES produziu grupos melhores atingindo 0,23 como o melhor valor de *recall*, enquanto o RSA obteve o valor 0,08 como valor máximo para a *recall*.

6.2.2 SEGUNDO CONJUNTO DE EXPERIMENTOS – CRIPTOGRAMAS COM TAMANHOS DISTINTOS

Este experimento tem o objetivo de tratar uma situação mais próxima da realidade, onde os tamanhos dos criptogramas podem variar, juntamente com a chave utilizada na cifragem. Assim, foram utilizados 300 criptogramas, de 10 tamanhos diferentes, gerados com o algoritmo DES, com 20 chaves distintas e aleatórias, com um total de 15 textos para cada chave (figura 6.7). Os tamanhos dos criptogramas em bytes são: 64, 128, 192, 256, 384, 512, 1024, 1536, 2048 e 3072.

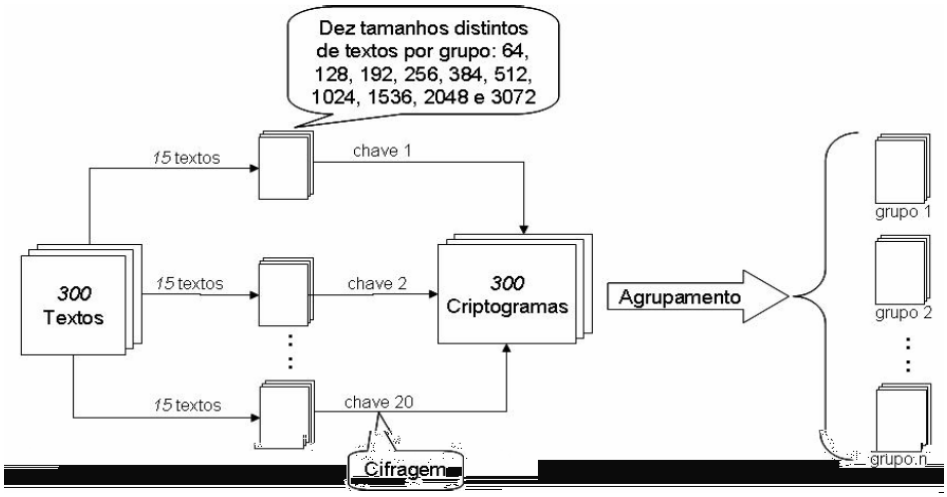


FIG. 6.7: Descrição do segundo conjunto de experimentos (CARVALHO, 2006)

A árvore de experimentos para este conjunto é semelhante a apresentada na figura 6.4, para o primeiro conjunto de experimentos, subconjunto para o algoritmo DES.

Um dos objetivos deste experimento é verificar a formação dos grupos, considerando que criptogramas com tamanhos que obtiveram valores de *recall* menores que um, 64 bytes, por exemplo, são misturados com tamanhos que obtiveram valores maiores, tendo em alguns casos, alcançado o valor máximo de *recall*. Na tabela 6.3, verifica-se o resultado obtido em Carvalho, o qual não teve valores máximos de *recall*, apenas para duas chaves. Entretanto, para 15 chaves, esse valor foi muito próximo de um, o que indica que a influência dos textos que isoladamente obtiveram baixo valor de *recall* é pequena.

TAB. 6.3: Resultados obtidos em Carvalho (2006) para o experimento segundo conjunto de experimentos

Abrangência
0.66 em 3 das chaves
0.8 em 2 das chaves
0.87 em 3 das chaves
0.93 em 10 das chaves
1 em 2 das chaves

O outro objetivo é verificar se os métodos de agrupamento e as medidas de similaridade e distância, utilizadas neste trabalho, produzem resultados melhores do que os obtidos por Carvalho (tabela 6.3).

6.2.2.1 RESULTADOS E AVALIAÇÕES

No apêndice 12, pode-se verificar que os resultados foram compatíveis com os relatados na tabela 6.3, não havendo melhora nos valores de *recall*, em função do método de agrupamento ou da medida de similaridade utilizada. Por outro lado, os resultados ficaram piores para as medidas de distância, de uma maneira geral, e para as medidas de similaridade no método *complete-link*.

Já os valores de *precision* foram um em todos os métodos para todas as medidas de similaridades e para a distância Bray-Curtis. Para as demais distâncias, os valores ficaram em 0,05, o que indica a formação de um único grupo composto por todos os criptogramas, sendo a única exceção a distância Manhattan, a qual obteve *precision* um, para 19 chaves, no método *single-link*.

Destes resultados, conclui-se, mais uma vez, que o método *single-link* trabalha melhor com as medidas de similaridade, o método *complete-link* trabalha melhor com a medida de distância Bray-Curtis, e o método *group average-link*, pode lidar bem tanto com as medidas de similaridade, quanto com a distância Bray-Curtis. As distâncias Euclidiana, Manhattan e Canberra não foram adequadas para o uso com os criptogramas neste experimento, já que os valores da *precision* foram baixos.

6.2.3 TERCEIRO CONJUNTO DE EXPERIMENTOS – SIMULAÇÃO DE CAPTURA DE CRIPTOGRAMAS

Neste experimento é realizada a simulação de uma captura de criptogramas. Assim, uma entidade externa preparou dois conjuntos de 20 textos em claro, no tamanho de 10 Kbytes cada, cifrados com 10 chaves distintas, com um total de 200 criptogramas em cada conjunto. As informações conhecidas sobre os dois conjuntos é que ambos foram cifrados no modo de operação ECB e que no primeiro conjunto os blocos possuíam 64 bits e no segundo conjunto os blocos possuíam 128 bits.

Um terceiro conjunto também foi disponibilizado, o qual possuía as mesmas características que os conjuntos anteriores, entretanto sendo cifrado no modo de operação CBC, com blocos de 128 bits. A árvore de experimentos pode ser vista na figura 6.8.

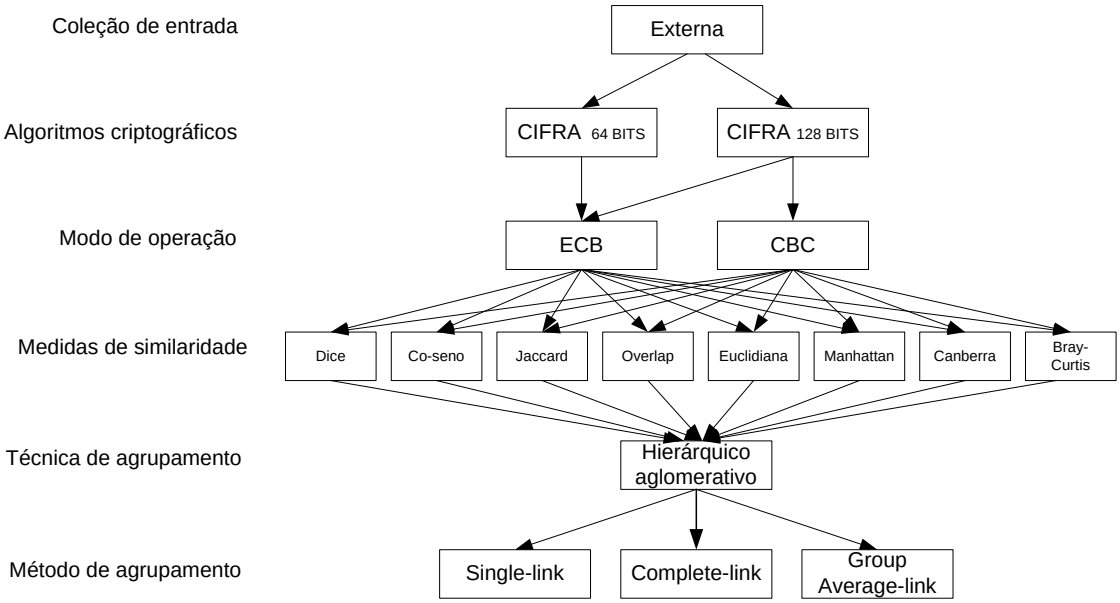


FIG. 6.8: Árvore de experimentos para o terceiro conjunto de experimentos

6.2.3.1 RESULTADOS E AVALIAÇÕES

No apêndice 13, estão descritos os resultados para este conjunto de experimentos. Mais uma vez, os resultados foram semelhantes aos obtidos por Carvalho não havendo melhora nos valores de *recall*, em função do método de agrupamento ou da medida de similaridade ou distância utilizada.

Na cifra de 64 *bits*, as medidas de similaridade alcançaram valores máximos de *precision* e *recall* em todos os métodos. De uma maneira geral, as medidas de distância se mostraram inadequadas. As exceções são a distância Manhattan, a qual obteve esses valores máximos no método *single-link*, e a distância Bray-Curtis que obteve esses valores no método *complete-link*, e o valor de *precision* um e *recall* 0,75 para os demais métodos.

Na cifra de 128 *bits*, no modo de operação ECB, as medidas de similaridade alcançaram valores máximos de *precision* e *recall*, com os métodos *single-link* e *group average-link*. Com o método *complete-link*, a *recall* ficou em 0,45. De uma maneira geral, as medidas de distância se mostraram mais uma vez inadequadas. A distância Bray-Curtis foi a exceção

obtendo maior *recall* do que as medidas de similaridades no método *complete-link*, mas com *recall* inferior a estas medidas para os demais métodos.

Já os valores de *precision* foram um em todos os métodos para todas as medidas de similaridades e para a distância Bray-Curtis. Já a distância Manhattan obteve valor um em alguns casos e as distâncias Canberra e Euclidiana não alcançaram o valor um.

Note que esses resultados são bastante parecidos aos obtidos nos experimentos com o DES (chave de 64 *bits*) e com o AES (chave de 128 *bits*), o que pode sugerir a classificação do algoritmo por meio do valor de *precision* e *recall*.

Na cifra de 128 *bits*, no modo de operação CBC, não foi possível realizar o agrupamento de criptogramas de acordo com a chave, já que, como visto anteriormente, a saída do processo de cifragem de um bloco é utilizada na cifragem do bloco seguinte, o que provavelmente elimina a repetição dos blocos.

Apesar disto, na tabela 6.4 pode-se perceber um grau de associação que, embora pequeno, questiona a segurança esperada para este modo de operação (SCHNEIER, 1996) e (MENEZES, 1996). Ressalta-se que os criptogramas agrupados na tabela 6.4 são decorrentes de textos diferentes. Estas associações foram obtidas como uso das medidas de similaridades e em todos os métodos de agrupamento.

Como visto na seção 2.3.1.2, blocos de criptogramas iguais não deveriam ser produzidos a partir de chaves diferentes, ainda mais com textos diferentes.

TAB. 6.4: Resultados obtidos para o agrupamento com cifra de 128 *bits*, no modo CBC

Chaves	Quantidade de grupos	Criptogramas por grupo
01 e 03	1	2
01 e 08	1	2
02 e 05	2	2
03 e 04	1	2
03 e 08	1	2
03 e 09	1	2
04 e 08	1	2
09 e 10	1	2

Embora, a primeira vista, a associação apresentada na tabela pareça formada ao acaso, uma análise mais atenta sugere que alguma característica presente na chave provoca o agrupamento, já que as chaves 02 e 05 formaram dois grupos, com um total de quatro

criptogramas, sendo estes criptogramas originados de quatro textos em claro diferentes. Ainda neste contexto, existem associações entre as chaves 01, 03, 04 e 08, conforme a figura 6.9 e a tabela 6.4. Ressalta-se que todos os criptogramas relacionados à figura 6.9 foram cifrados a partir de textos diferentes.

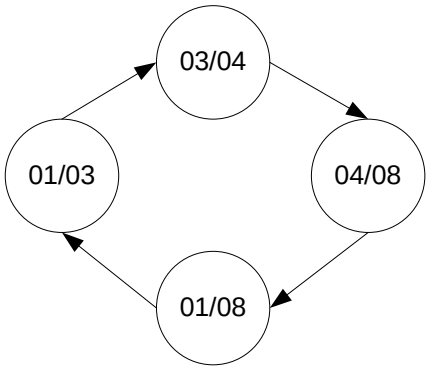


FIG. 6.9: Relação encontrada entre as chaves no modo CBC, conforme a TAB. 6.4

A questão que surge é: qual o grau de semelhança entre as chaves utilizadas na cifragem destes criptogramas? Neste experimento, não foi possível o acesso as chaves, uma vez que o experimento simula uma captura de criptogramas e a sua finalidade era testar o agrupamento. Mas, como pôde ser visto, outros níveis de associação, os quais não eram esperados, foram encontrados.

A tabela 6.4 mostra ainda que o valor de *precision* não foi um, mesmo com as medidas de similaridade, diferente do que era esperado, considerando o informado no trabalho de Carvalho (2006, 70).

6.2.4 QUARTO CONJUNTO DE EXPERIMENTOS – SEPARAÇÃO DE CRIPTOGRAMAS DE ACORDO COM O TAMANHO DA CHAVE

Este conjunto de experimentos tem a finalidade de verificar a possibilidade de agrupar criptogramas de acordo com o tamanho da chave utilizada na cifragem. O algoritmo AES, com os tamanhos de chaves de 128, 192 e 256 *bits* (150 chaves aleatórias e distintas), foi o escolhido para esse conjunto. Foram utilizados os mesmos conjuntos de criptogramas

descritos na seção 6.2.1.2. Em cada experimento foram utilizados 4.500 criptogramas, onde cada tamanho de chave contribuiu com 1.500 criptogramas.

Como visto nos experimentos anteriores, cada tamanho de chave tem o seu próprio tamanho de bloco. Uma vez que os criptogramas cifrados com tamanhos de chaves diferentes foram colocados juntos para passar pelo processo de agrupamento, tornou-se necessário testar os três tamanhos de blocos recomendados: 16, 24 e 32 *bits*. Pelos resultados apresentados anteriormente, o método *single-link* e o coeficiente Dice foram utilizados neste conjunto, cuja árvore de experimentos, pode ser vista na figura 6.10.

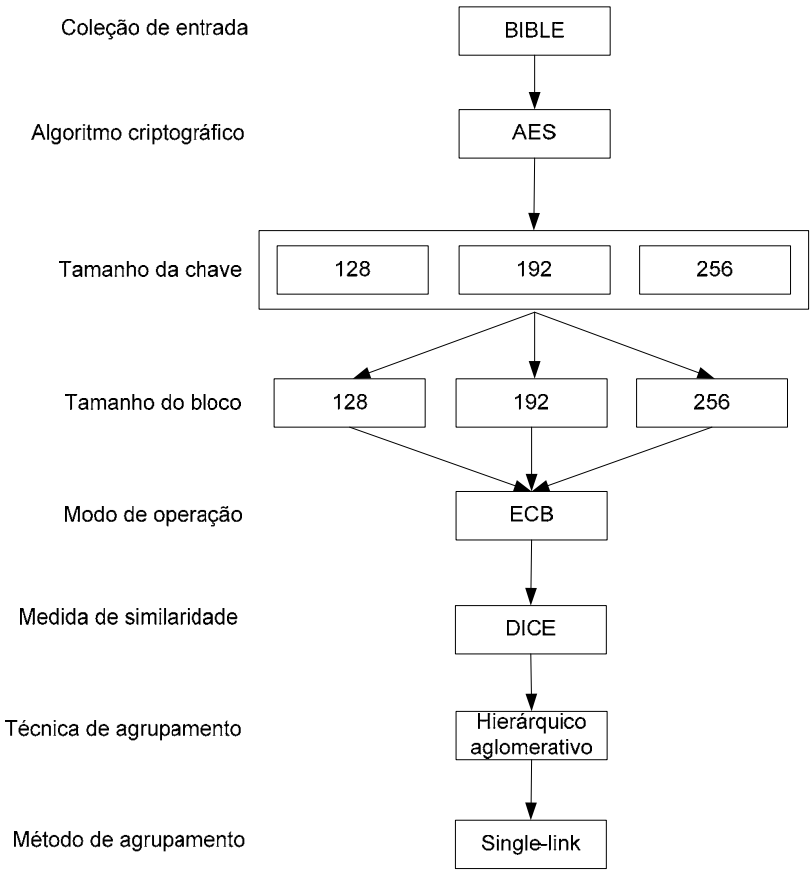


FIG. 6.10: Árvore de experimentos para o quarto conjunto

Os resultados estão descritos nas tabelas 6.5 e 6.6, onde se pode notar que o valor de *precision* foi um em todos os casos. Já a *recall* obteve os melhores resultados com blocos de 16 *bits*, sendo 0,02 o valor máximo alcançado.

TAB. 6.5: Resultados obtidos para quarto conjunto de experimentos

Tamanho do bloco (bits)	16		24		32	
Tamanho dos textos (bytes)	P	R	P	R	P	R
10240	1	0,02	1	0,003	1	0,005
8192	1	0,02	1	0,003	1	0,003
6144	1	0,02	1	0,002	1	0,002
4096	1	0,019	1	0,001	1	0,001
3072	1	0,017	1	0,001	1	0,001
2560	1	0,01	1	0,001	1	0,001
2048	1	0,007	1	0,001	1	0,001
1536	1	0,004	1	0,001	1	0,001
1024	1	0,003	1	0,001	1	0,001

TAB. 6.6: Grupos formados para quarto conjunto de experimentos

Tamanho do bloco (bits)	16	24	32
Tamanho dos textos (bytes)	Quantidade de grupos formados		
10240	150	3300	2550
8192	150	3300	3000
6144	150	4050	3600
4096	300	4350	4500
3072	600	4350	4500
2560	900	4350	4500
2048	1950	4350	4500
1536	3150	4500	4500
1024	3300	4500	4500

Observando os valores de *recall* para os criptogramas de 6144, 8192 e 10240 *bytes*, nota-se que a quantidade de grupos formados foi 150 e que cada grupo juntou 30 criptogramas, sendo que todos os 30 criptogramas em cada um dos grupos foram cifrados com a mesma chave. Uma vez que este conjunto de experimentos utiliza 150 chaves distintas, percebe-se que o agrupamento foi realizado com sucesso por chave, com *precision e recall máxima*.

Isto mostra que a chave utilizada na cifragem tem forte influência no processo de agrupamento de criptogramas e que o simples fato de a chave ter um tamanho igual, aparentemente, não gera características nos criptogramas que sejam passíveis de associações.

De fato, para que esses experimentos obtivessem os resultados esperados, o tamanho da chave deveria gerar características de associação nos criptogramas mais fortes do que as geradas por uma chave em particular, de modo que um grupo pudesse ser formado apenas por criptogramas cifrados com um mesmo tamanho.

6.2.5 QUINTO CONJUNTO DE EXPERIMENTOS – TENTATIVA DE IDENTIFICAÇÃO DO ALGORITMO CRIPTOGRÁFICO POR MEIO DOS CRIPTOGRAMAS GERADOS POR ESTES ALGORITMOS

Este conjunto de experimentos tem a finalidade de verificar a possibilidade de agrupar criptogramas de acordo com o algoritmo criptográfico utilizado na cifragem. Foram utilizados os algoritmos DES, AES e RSA, com os tamanhos de chaves de 64, 128, 512 *bits*, respectivamente (150 chaves aleatórias e distintas). Foram utilizados subconjuntos dos conjuntos de criptogramas descritos nas seções 6.2.1.1, 6.2.1.2 e 6.2.1.3. Nesses subconjuntos foram consideradas as interseções de tamanhos de textos e descartados os textos de 8192 e 10240, pois estes tamanhos consumiam recursos computacionais além dos disponíveis e, como visto no conjunto experimentos anterior, que é semelhante a este, a formação dos grupos se deu da mesma forma para os tamanhos 6144, 8192 e 10240.

Assim, os tamanhos selecionados foram: 1024, 2048, 4096 e 6144. Em cada experimento foram utilizados 4.500 criptogramas, onde cada algoritmo criptográfico contribuiu com 1.500 criptogramas.

Neste conjunto são considerados três tamanhos de blocos: 8, 16 e 64 *bits*, para os algoritmos DES, AES e RSA, respectivamente. O método *single-link* e o coeficiente Dice foram utilizados neste conjunto, cuja árvore de experimentos, pode ser vista na figura 6.11.

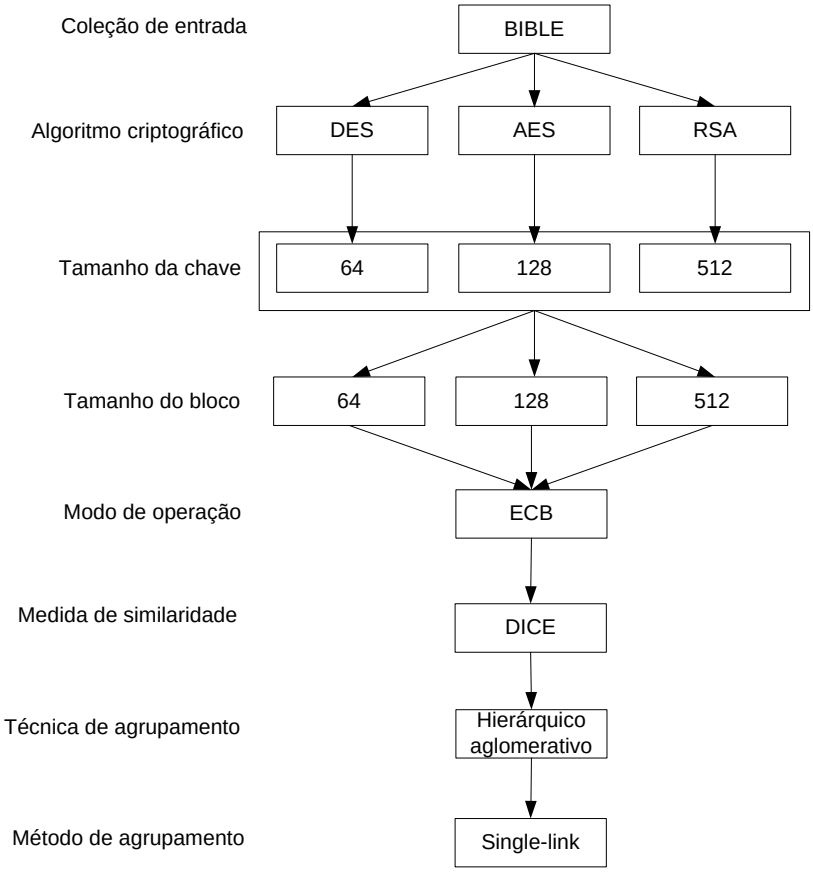


FIG. 6.11: Árvore de experimentos para o quinto conjunto

Os resultados podem ser vistos nas tabelas 6.7 e 6.8. Das tabelas, nota-se que o valor de *precision* foi um em todos os casos. Já a *recall* obteve os melhores resultados tanto com blocos de 8 *bits* quanto com blocos de 16 *bits*, sendo 0,02 o valor máximo alcançado. Os grupos formados para esses dois tamanhos de blocos também foram semelhantes, mas para textos menores os blocos de 8 *bits* formaram menos grupos o que significa que mais criptogramas ficaram juntos em menos grupos, já que a *precision* foi um.

TAB. 6.7: Resultados obtidos para quinto conjunto de experimentos

Tamanho do bloco (bits)	8		16		64	
Tamanho dos textos (bytes)	P	R	P	R	P	R
6144	1	0,02	1	0,02	1	0,001
4096	1	0,019	1	0,02	1	0,001
2048	1	0,007	1	0,007	1	0,001
1024	1	0,003	1	0,003	1	0,001

TAB. 6.8: Grupos formados para quinto conjunto de experimentos

Tamanho do bloco (bits)	8	16	64
Tamanho dos textos (bytes)	Quantidade de grupos formados		
6144	1600	1600	4500
4096	1200	1250	4050
2048	2200	2800	4500
1024	2650	3700	4500

Por meio de uma análise dos resultados com os melhores valores de *recall* e com a menor quantidade de grupos formados (textos com 6144 *bytes*), nota-se que a quantidade de grupos formados foi 50 para o AES e 50 grupos para o DES e que cada grupo juntou 30 criptogramas, sendo que todos os 30 criptogramas em cada um dos grupos foram cifrados com a mesma chave. Já para o RSA foram formados 1.500. Observando os experimentos anteriores, verifica-se que os resultados obtidos são idênticos aos obtidos neste conjunto de experimentos, confirmando a influência da chave utilizada na cifragem no processo de agrupamento de criptogramas.

Com as técnicas utilizadas, o algoritmo criptográfico, aparentemente, não gera características de associação nos criptogramas mais fortes do que as geradas por uma chave em particular, a ponto de um grupo ser formado apenas por criptogramas cifrados com um mesmo algoritmo criptográfico.

6.2.6 SEXTO CONJUNTO DE EXPERIMENTOS – INFLUÊNCIA DO TAMANHO DO CRIPTOGRAMA COM TEXTOS MAIORES

Dos experimentos realizados, pode ser visto que para o algoritmo AES, com chaves a partir de 192 *bits*, e para o algoritmo RSA, com chaves a partir de 256 *bits*, o agrupamento não atingiu *precision e recall máxima*. Assim, este experimento considerará chaves de 192 *bits* e 256 *bits* para o AES e de 256 *bits* para o RSA.

Uma vez que à medida que os textos aumentam de tamanho, também melhora o resultado do agrupamento, este conjunto tem a finalidade de verificar o agrupamento com

textos de tamanhos maiores do que 10240 *bytes*. A tabela 6.9 apresenta uma relação aproximada entre os tamanhos de textos e a quantidade de páginas[28] que eles possuem.

TAB. 6.9: Quantidade de páginas por tamanho de texto

Tamanho dos textos	Quantidade aproximada de páginas
20160	7
18592	6
16192	5
14592	5
12240	4
10240	4
8192	3
6144	2

6.2.6.1 SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO AES

Foram definidos cinco tamanhos de textos diferentes, em bytes: 12240, 14592, 16192, 18592 e 20160, cada tamanho com 30 textos, cifrados com 50 chaves diferentes e aleatórias, utilizados dois tamanhos diferentes de chaves (100 chaves aleatórias), totalizando 15.000 criptogramas. Na figura 6.12, pode ser observado uma árvore que descreve os experimentos realizados.

[28] Deve-se considerar que o tipo e o tamanho da fonte influenciam na quantidade de páginas.

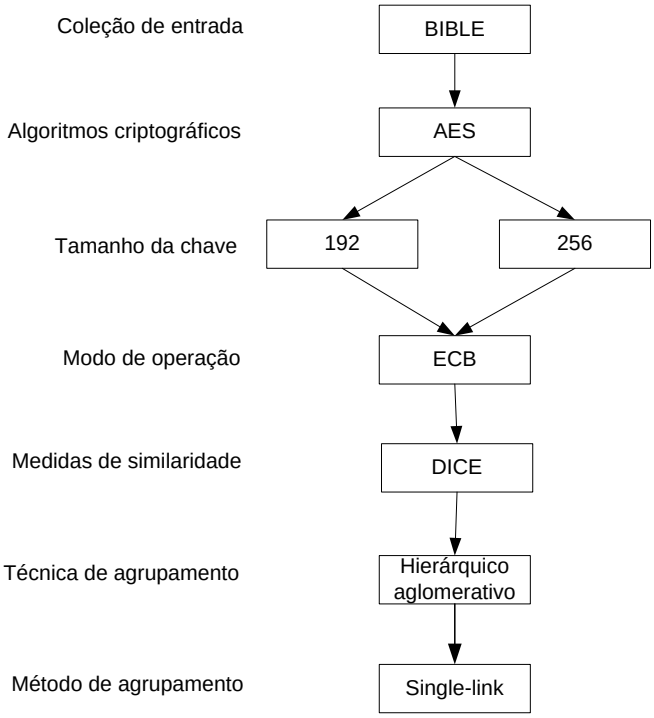
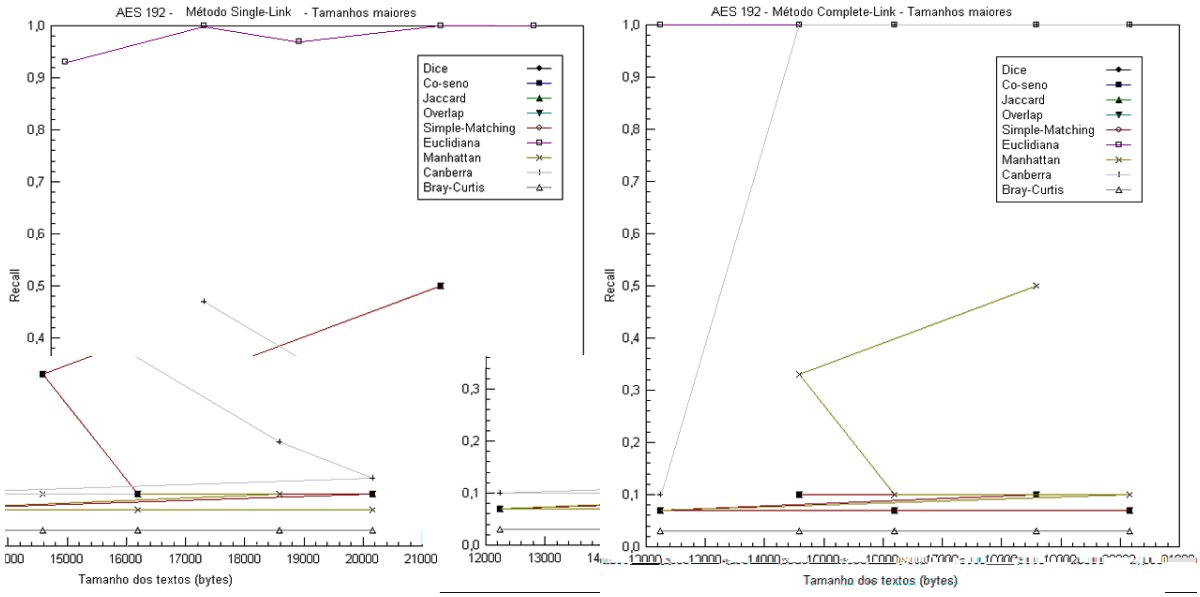


FIG. 6.12: Sexto conjunto de experimentos: subconjunto para o algoritmo AES

6.2.6.1.1 RESULTADOS E AVALIAÇÕES

Os resultados para este subconjunto podem ser visualizados nos gráficos 6.17 e 6.18, (a) e (b), e nos apêndices 14 e 15. Como pode ser observado nos gráficos e nos apêndices, os resultados mostraram que houve melhora no valor de *recall*, o qual em alguns casos foi até três vezes maior. Nenhuma das medidas alcançou *precision e recall máxima*. Embora a mera observação dos gráficos possa indicar que as medidas de distância tiveram valores altos de *recall*, consultando os apêndices pode-se ver que nestes casos a precisão ficou comprometida.

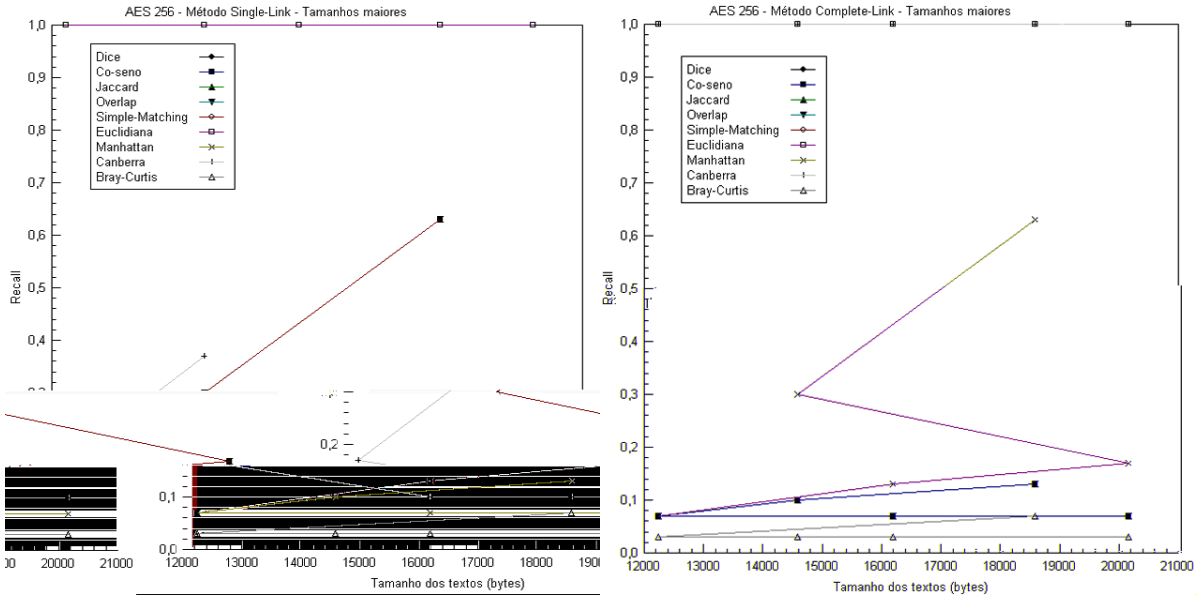
Diferente do que era esperado, o segundo maior tamanho de texto (18592 *bytes*) obteve recall maior do que o maior tamanho de texto (20160 *bytes*) nos dois tamanhos de chaves. Ainda neste caso, outro item inesperado foi o fato de o maior tamanho de chave (256 *bits*) obteve 0,63 de recall, enquanto que para o tamanho de 192 *bits* o valor foi de 0,50. Esses valores foram obtidos com as medidas de similaridade, nos métodos *single-link* e *group average-link*, e com a distância Manhattan no método *complete-link*.



(a) (b)

GRA. 6.17: *Recall* para o algoritmo AES, com chaves de 192 *bits*, com os métodos *Single-Link* e *Complete-Link*

Os valores de *precision* foram um para todas as medidas de similaridade e para as distâncias Manhattan e Bray-Curtis, em todos os métodos de agrupamento. As demais medidas distância alcançaram um somente em alguns casos.



(a) (b)

GRA. 6.18: *Recall* para o algoritmo AES, com chaves de 256 *bits*, com os métodos *Single-Link* e *Complete-Link*

6.2.6.2 SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO RSA

Foram definidos cinco tamanhos de textos diferentes, em bytes: 12240, 14600, 16200, 18600 e 20160, cada tamanho com 30 textos, cifrados com 50 chaves diferentes e aleatórias, totalizando 7.500 criptogramas. Na figura 6.13, pode ser observado uma árvore que descreve os experimentos realizados.

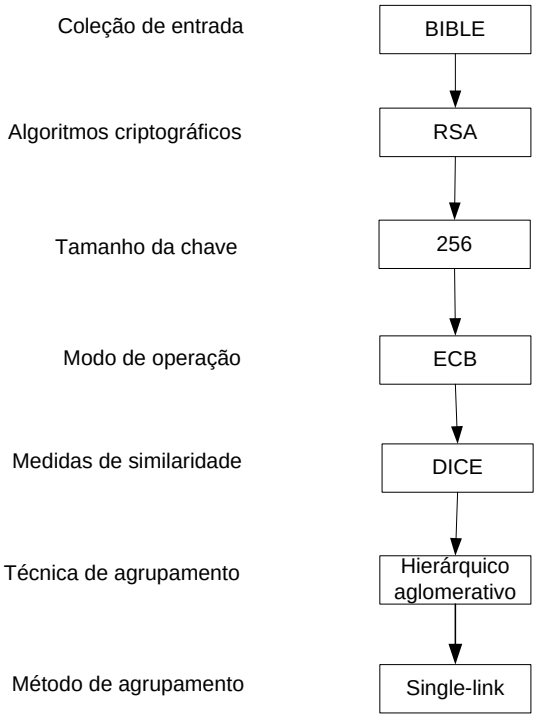


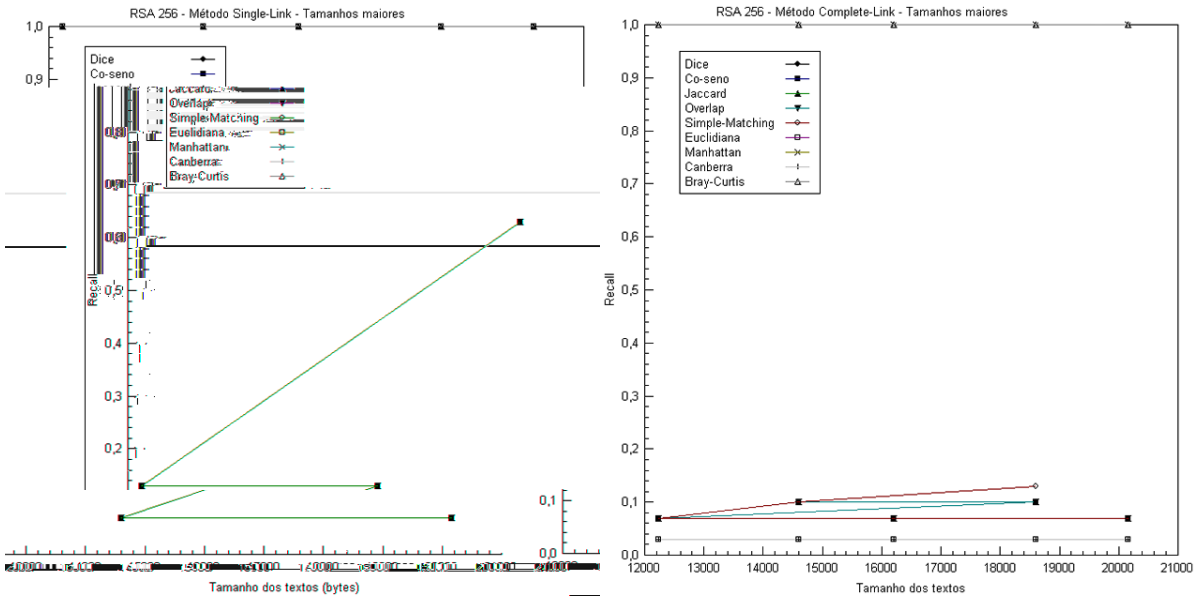
FIG. 6.13: Sexto conjunto de experimentos: subconjunto para o algoritmo RSA

6.2.6.2.1 RESULTADOS E AVALIAÇÕES

Os resultados para este subconjunto são apresentados nos gráficos 6.19, (a) e (b), e no apêndice 16. Como pode ser observado nos gráficos e no apêndice, os resultados mostraram que houve melhora no valor de *recall*, o qual em alguns casos foi até três vezes maior. Mas, Nenhuma das medidas alcançou *precision e recall máxima*.

Assim como no conjunto anterior, apenas observação dos gráficos indicará que as medidas de distância tiveram valores altos de *recall*, mas consultando os apêndices será visualizado o comprometimento da precisão.

Diferente do que era esperado, o segundo maior tamanho de texto (18592 *bytes*) obteve 0,63 de *recall* enquanto o maior tamanho de texto (20160 *bytes*) obteve 0,07 para esse valor. Esses valores foram obtidos com as medidas de similaridade, nos métodos *single-link* e *group average-link*. Para as medidas de distância, os valores de *recall* indicam a formação de apenas um grupo em alguns casos e a formação de grupos com apenas um criptograma cada.



(a) (b)

GRA. 6.19: *Recall* para o algoritmo RSA, com chaves de 256 *bits*, com os métodos *Single-Link* e *Complete-Link*

Os valores de *precision* foram um para todas as medidas de similaridade. Já as medidas distância alcançaram um somente em alguns casos.

6.2.7 SÉTIMO CONJUNTO DE EXPERIMENTOS – CLASSIFICAÇÃO DE CHAVES POR MEIO DE UMA REDE NEURAL ARTIFICIAL

6.2.7.1 SUBCONJUNTO DE EXPERIMENTOS COM CRIPTOGRAMAS DE 2048 BYTES

Para esse experimento foram utilizados 150 criptogramas, cifrados com cinco chaves diferentes, cada uma cifrando 30 textos. Destes criptogramas, 80% (120 criptogramas) foram

utilizados na fase de treino (agrupamento) e os demais 20% (30 criptogramas) foram utilizados na fase de teste (classificação). As representações nos mapas foram feitas por meio de cores de acordo com a tabela 6.10. A rede foi configurada com tamanho 10 X 10, cinco épocas, taxa de aprendizado de 0,1 e medida do ângulo do co-seno.

TAB. 6.10: Representação das chaves no mapa bidimensional

Chave	Cor
1	Preto
2	Cinza-claro
3	Cinza
4	Vermelho
5	Azul

6.2.7.1.1 RESULTADOS

Na figura 6.14, está o resultado para o agrupamento (estágio de treino). Verifica-se que o agrupamento formou grupos topologicmante bem definidos para as chaves 2 e 3. Para as demais chaves, os criptogramas ficaram muito próximos, indicando que a separação não foi boa. O apêndice 19 apresenta o resultado completo para a fase de treino e teste.

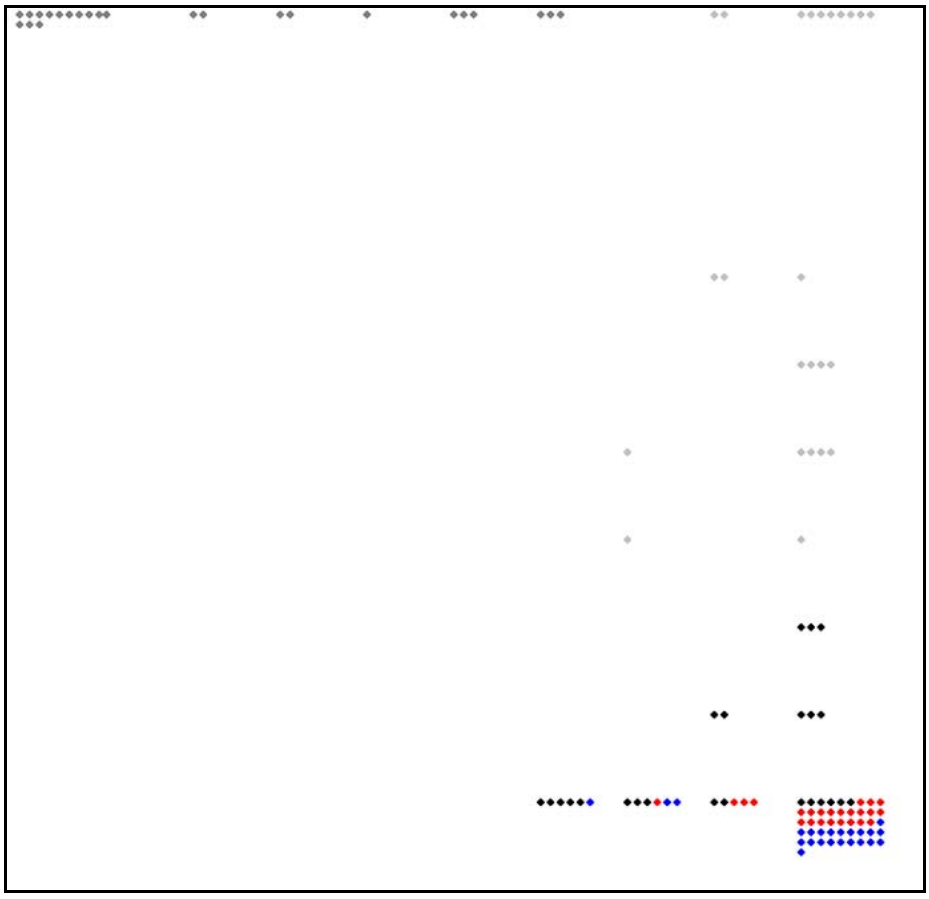


FIG. 6.14: Mapa formado no estágio de treino com criptogramas de 2048 *bytes*

Na figura 6.15, está o resultado para a classificação (estágio de teste). Embora o resultado do agrupamento não tenha sido bom, a classificação foi realizada de maneira adequada, isto é, criptogramas diferentes cifrados com a mesma chave foram mapeados no mesmo neurônio ou em neurônios próximos, tanto na fase de agrupamento, quanto na fase de classificação. Isto caracteriza o sucesso da classificação.

Na fase de treino a dimensão do espaço vetorial ficou em 26.535 dimensões e o tempo de execução foi de 4 horas e 20 minutos. Na fase de teste, a dimensão do espaço vetorial ficou em 7.170 dimensões e o tempo de execução foi de 25 minutos. Estes resultados estão coerentes com o estudo apresentado no apêndice 17.

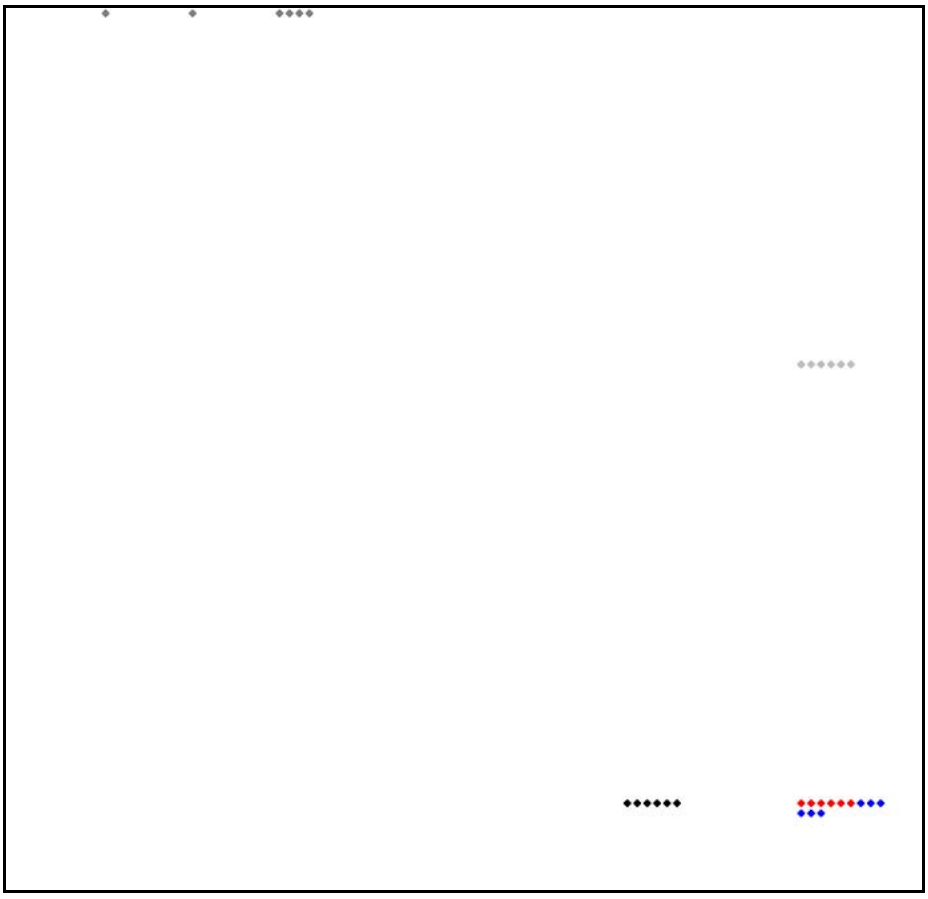


FIG. 6.15: Mapa formado no estágio de teste com criptogramas de 2048 *bytes*

6.2.7.2 SUBCONJUNTO DE EXPERIMENTOS COM CRIPTOGRAMAS DE 6144 BYTES

Uma vez que os resultados do experimento anterior não foram bons, do ponto de visto do agrupamento, este novo experimento foi definido com criptogramas de tamanhos maiores para verificar se a fase de agrupamento pode alcançar melhor resultado. Assim, foram utilizados 300 criptogramas, cifrados com dez chaves diferentes, cada uma cifrando 30 textos. O tamanho dos criptogramas foi de 6144 *bytes*. Destes criptogramas, 80% (240 criptogramas) foram utilizados na fase de treino (agrupamento) e os demais 20% (60 criptogramas) foram utilizados na fase de teste (classificação). As representações nos mapas foram feitas por meio de cores de acordo com a tabela 6.11. A rede foi configura com tamanho 10 X 10, cinco épocas, taxa de aprendizado de 0,1 e medida do ângulo do co-seno.

TAB. 6.11: Representação das chaves no mapa bidimensional

Chave	Cor
1	Preto
2	Cinza-claro
3	Cinza
4	Vermelho
5	Azul
6	Cyan
7	Verde
8	Magenta
9	Laranja
10	Amarelo

6.2.7.2.1 RESULTADOS

Na figura 6.16, está o resultado para o agrupamento (estágio de treino). Verifica-se que o agrupamento formou grupos topologicmante bem definidos para as chaves 3, 5 e 8. Para as demais chaves, os criptogramas ficaram muito próximos, indicando que a separação não foi boa. O apêndice 20 apresenta o resultado completo para a fase de treino e teste.

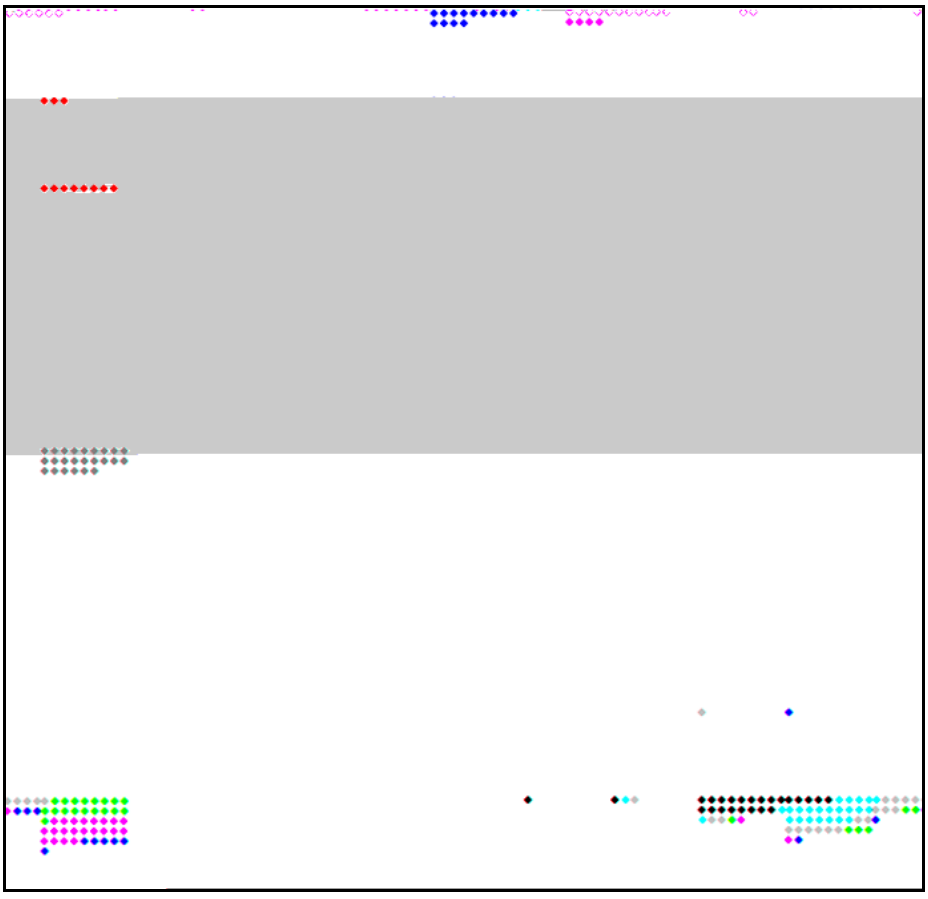


FIG. 6.16: Mapa formado no estágio de treino com criptogramas de 6144 *bytes*

Na figura 6.17, está o resultado para a classificação (estágio de teste). Da mesma forma que o experimento anterior, Embora o resultado do agrupamento não tenha sido bom, a classificação foi realizada de maneira adequada, isto é, criptogramas diferentes cifrados com a mesma chave foram mapeados no mesmo neurônio ou em neurônios próximos, tanto na fase de agrupamento, quanto na fase de classificação. Isto caracteriza o sucesso da classificação.

Na fase de treino a dimensão do espaço vetorial ficou em 138.430 dimensões e o tempo de execução foi de 12 horas e 40 minutos. Na fase de teste, a dimensão do espaço vetorial ficou em 41.367 dimensões e o tempo de execução foi de 55 minutos. Estes resultados estão coerentes com o estudo apresentado no apêndice 17.

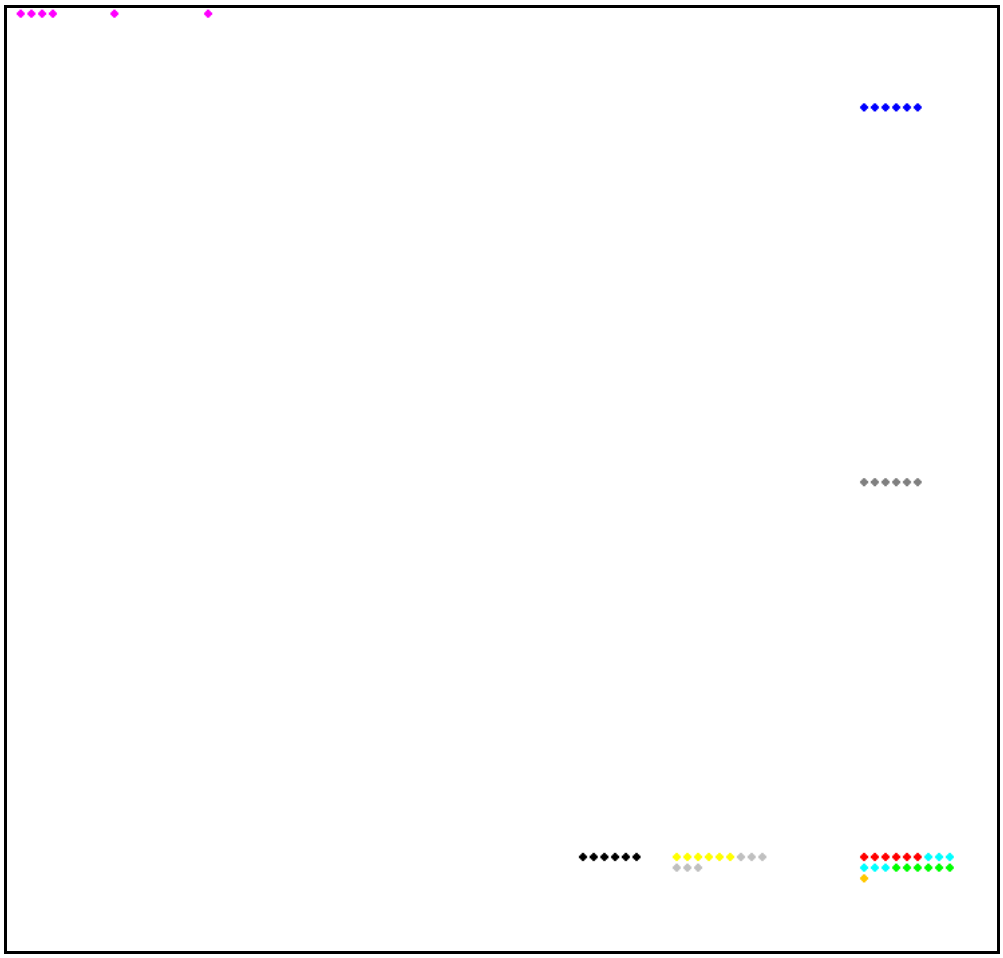


FIG. 6.17: Mapa formado no estágio de teste com criptogramas de 6144 *bytes*

6.6 AVALIAÇÃO ESTATÍSTICA DAS MEDIDAS DE SIMILARIDADE E DISTÂNCIA

A tabela 6.11 apresenta uma matriz simétrica, a qual mostra o resultado da aplicação do Coeficiente de Correlação de Pearson sobre as medidas de similaridade e distância. Percebe-se que todas as medidas estão linearmente relacionadas, já que os valores obtidos estão muito próximos de um e em alguns casos são exatamente um, como Co-seno, Dice e Overlap e Bray-Curtis e Manhattan.

Destes resultados, poderia se esperar que os resultados dos agrupamentos fossem semelhantes em todos os casos. Entretanto, os experimentos mostraram que apenas a correlação linear não é suficiente para avaliar as medidas. Além disso, o critério de parada também influenciou na maneira como os grupos foram formados.

TAB. 6.12: Coeficiente de Correlação de Pearson para as medidas utilizadas

	Dice	**Co-seno**	**Jaccard**	**Overlap**	**Euclidiana**	**Manhattan**	**Bray-Curtis**	**Canberra**
Dice	1	1	0,99995	1	0,99993	0,99996	0,99996	0,99991
Co-seno	1	1	0,99995	1	0,99993	0,99996	0,99996	0,99991
Jaccard	0,99995	0,99995	1	0,99995	0,99996	0,99999	0,99999	0,99997
Overlap	1	1	0,99995	1	0,99993	0,99996	0,99996	0,99990
Euclidiana	0,99993	0,99993	0,99996	0,99993	1	0,99996	0,99996	0,99988
Manhattan	0,99996	0,99996	0,99999	0,99996	0,99996	1	1	0,99997
Bray-Curtis	0,99996	0,99996	0,99999	0,99996	0,99996	1	1	0,99997
Canberra	0,99991	0,99991	0,99997	0,99990	0,99988	0,99997	0,99997	1

7 CONCLUSÕES E CONSIDERAÇÕES FINAIS

A criptoanálise efetiva dos sistemas criptográficos atuais parece uma perspectiva distante. Entretanto, a identificação de padrões em criptogramas pode contribuir no processo de busca por fraquezas nestes sistemas. Além disso, a existência de padrões em criptogramas pode ser utilizada também como requisito para a avaliação de produtos criptográficos.

Considerando que a criptoanálise clássica era baseada nas características lingüísticas do idioma de origem do texto em claro, pode-se supor que as técnicas de classificação de textos obtenham algum sucesso na busca por padrões em criptogramas. De fato, em Carvalho (2006), observa-se a realização do agrupamento de criptogramas, de maneira que os grupos formados possuíam apenas criptogramas cifrados com uma mesma chave.

Neste contexto, a presente dissertação usou técnicas de agrupamento e classificação de textos, o algoritmo RSA e chaves maiores que 128 *bits* para o algoritmo AES, que não foram considerados no trabalho de Carvalho. Além disso, foi realizado estudo e aplicação de redes neurais artificiais com os objetivos de agrupamento e classificação de criptogramas.

Assim, além da medida do ângulo do co-seno e da técnica de agrupamento[29] *single-link*, foram usadas quatro outras medidas de similaridade, por meio dos coeficientes: Dice, Overlap, Jaccard e Simple-matching; quatro outras medidas de dissimilaridade, por meio das distâncias: Euclidiana, Manhattan, Canberra e Bray-Curtis; duas outras técnicas de agrupamento: *complete-link* e *group average-link*.; o mapa de Kohonen foi utilizado em experimentos de agrupamento e classificação de chaves e de algoritmos criptográficos; O coeficiente de Pearson foi utilizados para avaliar a medidas de similaridade e distância. Essas tecnologias foram viabilizadas por meio do desenvolvimento de módulos e ferramentas de software, conforme descrito no capítulo 5.

Todas as técnicas descritas acima foram aplicadas ao algoritmo DES, chave de 64 *bits*, algoritmo AES, chaves de 128, 192 e 256 *bits* e algoritmo RSA, chaves de 64, 128, 256, 512 e 1024 *bits*.

As medidas de similaridades se mostraram adequadas ao uso com criptogramas, não havendo grande diferença na contribuição de cada uma delas ao processo de agrupamento e

[29] Neste capítulo quando for dito somente agrupamento, queira considerar o agrupamento hierárquico aglomerativo. Quando o agrupamento se referir a outra técnica será enfatizado a técnica no texto.

requerendo tempos de execução bem próximos uns dos outros, com todos os métodos de agrupamento considerados neste trabalho.

Já as medidas de distância apresentaram desempenho variável. A distância Manhattan obteve bons resultados, principalmente com o método *complete-link*. Tais resultados para a distância Manhattan sugerem que essa medida de distâncias pode ser equivalente às medidas de similaridades, quando são considerados os métodos de agrupamento.

Alguns resultados para a distância Bray-Curtis foram bons, entretanto dispersos pelos métodos de agrupamento, isto é, em alguns experimentos esta distância foi melhor com o método *complete-link* e em outros, foi melhor com o *single-link* ou *group average-link*. Em geral, a distância Euclidiana e a distância Canberra não obtiveram bons resultados em nenhum dos métodos de agrupamento.

Os tempos de execução para as medidas de distância foram muito maiores do que os tempos requeridos pelas medidas de similaridade. A distância Manhattan obteve desempenho razoável, considerando o resultado obtido.

Ainda quanto aos tempos de execução, dos resultados pode-se notar que os tempos de execução caíram quando o tamanho da chave aumentou. Uma vez que quanto maior a chave, maior o espaço de termos utilizados, seria esperado que os tempos de execução aumentassem. Contudo, deve-se considerar que o tamanho dos criptogramas influencia neste tempo. Por exemplo, dados dois criptogramas c_1 e c_2 de igual tamanho. Seja 10240 *bytes* este tamanho. $c_1 \in 2^{64}$, com blocos de 64 *bits*, e $c_2 \in 2^{128}$, com blocos de 128 *bits*. Então c_1 terá um total de 1280 blocos e c_2 terá um total de 640 blocos. Mais blocos aumentam a possibilidade de blocos diferentes, aumentando assim a dimensão do espaço vetorial, levando ao aumento do tempo de execução. Logo, justifica-se o maior tempo de execução para criptogramas cifrados com chaves menores em relação aos criptogramas cifrados com chaves maiores, quando estes têm tamanho igual.

O Coeficiente de Correlação Linear de Pearson foi calculado para as medidas de similaridade e distância, par a par, o que comprovou a forte correlação entre as mesmas, já que todos os pares de medidas calculados tiveram valores próximos a um, com alguns pares chegando ao valor um, que indica correlação total.

Os métodos de agrupamento foram avaliados por meio do cálculo de *precision* e *recall*. Este cálculo revelou que os métodos *single-link* e *group average-link* formam melhores grupos quando utilizados com as medidas de similaridades. Já o método *complete-link* com a

distância Manhattan tem resultados semelhantes aos obtidos pelas medidas de similaridades nos demais métodos de agrupamento utilizados neste trabalho.

Quantos aos algoritmos criptográficos, foi possível obter *precision e recall máxima* com tamanho de criptogramas de até 10240 *bytes*, para o algoritmo DES, chave de 64 *bits*, algoritmo AES, chave de 128 *bits* e algoritmo RSA, chaves de 64 e 128 *bits*.

Chaves maiores foram testadas com textos maiores (12240 até 20160 *bytes*). Algoritmo AES, chaves de 192 e 256 *bits* e algoritmo RSA, chaves de 256 *bits*. Os valores de *recall* indicaram a formação de grupos melhores, mas não foi possível obter *precision* e *recall máxima*.

Os experimentos para o agrupamento de criptogramas cifrados com o mesmo algoritmo, mas com chaves de tamanho diferente e para a identificação dos algoritmos criptográficos revelaram que, com as técnicas de agrupamento utilizadas, os tamanhos de chaves utilizados nos algoritmos criptográficos e os próprios algoritmos criptográficos utilizados para cifrar um texto, aparentemente, não geram características de associação nos criptogramas mais fortes do que as geradas por uma chave qualquer. Isto quer dizer que a linguagem é determinada por uma chave, com os blocos representando os termos dessa linguagem e que os algoritmos criptográficos, assim como os tamanhos de chaves utilizados com esses algoritmos, não têm, aparentemente, influência na composição dessa linguagem. Contudo, isso só pode ser afirmado quando o modo de detecção dessa influência são os métodos de agrupamento hierárquico aglomerativo que foram utilizados neste trabalho e as tecnologias associadas a esses métodos.

Um experimento com uma rede neural artificial, baseada no mapa de Kohonen, mostrou bom potencial de uso dessa tecnologia para a identificação de algoritmos criptográficos a partir dos criptogramas. Como resultado, foi gerado um mapa bidimensional que apresentou criptogramas gerados com o mesmo algoritmo criptográfico, mas com chaves diferentes, agrupados topologicamente próximos. Em algumas partes do mapa, criptogramas gerados por algoritmos criptográficos diferentes também foram agrupados próximos.

Outro experimento, utilizando essa rede neural, foi realizado para a classificação de chaves. O mapa formado apresentou grupos bem definidos para algumas chaves, mas juntou criptogramas cifrados com chaves diferentes no mesmo grupo. Embora o resultado do agrupamento não tenha sido bom, a classificação foi realizada de maneira adequada, isto é, criptogramas diferentes cifrados com a mesma chave foram mapeados no mesmo neurônio ou em neurônios próximos, tanto na fase de agrupamento, quanto na fase de classificação.

Conclui-se que o mapa de Kohonen tem grande potencial para o agrupamento e classificação de criptogramas independente do critério de associação (chave ou algoritmo criptográfico, por exemplo). Mas, é necessário aperfeiçoá-lo para que o mesmo tenha tempo de execução e consumo memória factíveis, conforme mostrado no apêndice 17.

Um estudo feito nesta dissertação sobre o aperfeiçoamento de um módulo para calcular o coeficiente de Spearman, revelou um caminho para a melhoria do desempenho da rede neural artificial. Nesse estudo, para uma base com 1.500 criptogramas, existiam 1.124.250 pares passíveis de possuírem similaridades. Contudo, apenas 1.900 desses pares possuíam similaridade maior que zero, considerando que uma medida de similaridade tenha sido usada para o cálculo das similaridades.

Neste trabalho, a maior parte dos experimentos utilizou criptogramas cifrados no modo de operação ECB. Os resultados obtidos confirmaram a vulnerabilidade desse modo de operação, a qual já foi comentada em Schneir (1996). Entretanto, apesar da segurança esperada para o modo de operação CBC (SCHNEIER, 1996), (MENEZES, 1996) e (CARVALHO, 2006), os experimentos revelaram determinado nível de relacionamento entre as chaves e os criptogramas, pela análise da tabela 6.4. Esses indícios indicam um potencial de estudo sobre esse modo de operação na classificação de chaves e algoritmos criptográficos.

7.1 TRABALHOS RELACIONADOS

Como apresentado no início desta dissertação, a pesquisa sobre identificação de padrões por meio de técnicas de classificação de textos é uma técnica quase inexplorada. Podemos citar apenas o trabalho de Carvalho (2006), o qual apresentou bons resultados com o agrupamento de criptogramas pela chave utilizada na cifragem.

7.2 CONTRIBUIÇÕES DESTE TRABALHO

a) Foi confirmada a vulnerabilidade do modo de operação ECB;
b) Foram identificados padrões em criptogramas cifrados com o modo de operação CBC;
c) Foi confirmada a forte dependência da chave utilizada na cifragem para o processo de agrupamento dos criptogramas;

d) Foi desenvolvida uma rede neural artificial, baseada no mapa de Kohonen, para o agrupamento e classificação de criptogramas;
e) Foi realizado um estudo parcial sobre as necessidades de memória e tempo de execução requerido para a rede neural artificial, baseada no mapa de Kohonen, que foi utilizada neste trabalho, realizar o agrupamento e classificação de criptogramas;
f) Foram criadas e ampliadas ferramentas para o agrupamento criptogramas, possibilitando a execução de um grande número de experimentos de maneira automática utilizando medidas e métodos de agrupamento diversos, combinados entre si;
g) Foi verificada a aplicação da rede neural, desenvolvida neste trabalho, para o agrupamento e classificação de criptogramas pela chave utilizada na cifragem;
h) Foi verificada a aplicação da rede neural, desenvolvida neste trabalho, para o agrupamento e classificação de criptogramas pelo algoritmo criptográfico;
i) Foi verificada a possibilidade de agrupamento de criptogramas por tamanho de chave, onde os mesmo tenham sido cifrados com o mesmo algoritmo, mas com chaves de tamanho diferente;
j) Foi verificada a possibilidade de agrupamento de criptogramas por algoritmo criptográfico, dado um determinado criptograma;
k) Foi verificada a resistência dos algoritmos criptográficos AES e RSA com chaves maiores que 192 bits, inclusive, para os tamanhos de textos de até 10240, considerando as técnicas utilizadas neste trabalho;
l) Foi verificado que calcular a precision e a recall baseada no grupo composto pela maior quantidade de criptogramas (melhor grupo) produz resultados mais regulares do que uma abordagem baseada estritamente na micro-média;
m) Foram identificadas medidas de similaridades e distância que podem ser usadas apropriadamente para o agrupamento de criptogramas, assim como os tempos de execução paras as medidas de similaridade e distância;
n) Foram identificados métodos de agrupamento associados às medidas de similaridade ou distância, que podem ser utilizados para o agrupamento de criptogramas para formar bons grupos; e
o) Foi desenvolvida uma interface para a submissão de criptogramas via internet.

7.3 TRABALHOS FUTUROS

a) Desenvolver um módulo para integrar todas as ferramentas desenvolvidas para o agrupamento e classificação dos criptogramas;
b) Realizar mais testes de agrupamento com os algoritmos criptográficos AES e RSA com chaves maiores que 192 *bits*, inclusive, para verificar com que tamanhos de criptogramas esses algoritmos alcançam *precision e recall máxima*;
c) Buscar novos padrões que possibilitem o agrupamento de criptogramas por tamanho de chave, onde os mesmo tenham sido cifrados com o mesmo algoritmo, mas com chaves de tamanho diferente;
d) Buscar novos padrões que possibilitem o agrupamento de criptogramas por algoritmo criptográfico, com a finalidade de identificar o algoritmo criptográfico, dado um determinado criptograma;
e) Aperfeiçoar a rede neural desenvolvida neste trabalho, para que a mesma realize o agrupamento e classificação de criptogramas com sucesso, independente do critério de associação; e
f) Aprofundar os estudos sobre criptogramas cifrados com o modo de operação CBC, buscando identificar características que possibilitem agrupar e classificar esses criptogramas, independente do critério de associação.

8 REFERÊNCIAS BIBLIOGRÁFICAS

BARROS NETO, B; SCARMINIO, I. S; BRUNS, R. E. **Como fazer experimentos:** pesquisa e desenvolvimento na ciência e na indústria. 2 ed. Campinas: Editora da UNICAMP, 2003.

BEUTELSPACHER, Albrecht. **Cryptology**. 2 ed. Washington: Mathematical Association of America, 1994.

BIBLE. **Bible in basic english**. Disponível: http://www.o-bible.com/bbe.html [capturado 13 dez. 2005].

BIHAM, Eli; Shamir Adi. **Differential cryptanalysis of DES-like cryptsystems**. Journal of Cryptology, 1991, 4(1): 3 – 72.

BRZUSTOWSKI, John. **Clustering Calculator**. 2002. Disponível: http://www2.biology.ualberta.ca/jbrzusto/cluster.php [capturado 13 nov. 2005].

CARVALHO, Carlos A. B. de. **O uso de técnicas de recuperação de informações em criptoanálise**. 2006. 75 f. Dissertação (Mestrado em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro.

CHAPMAN, Sam. **String metrics**. 2006. Disponível: http://www.dcs.shef.ac.uk/~sam/stringmetrics.html [capturado 09 ago. 2006].

COELHO, Leandro de A. L. **Busca padrões em criptogramas:** um estudo de caso com o algoritmo RSA. 2006. 63 f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação). Orientação de William Augusto Rodrigues de Souza. Instituto Metodista Bennett, Rio de Janeiro.

COUTINHO, S. C. **Números inteiros e criptografia RSA**. Rio de Janeiro: IMPA/SBM, 2000.

FALOUTSOS, Christos. Signature files. In: FRAKES, William B, YATES, Ricardo B. **Information retrieval**: data structures and algorithms. Upper Saddle River: Prentice Hall, 1992. p. 363-392.

FAUSSET, Laurene. **Fundamentals of neural networks**: architectures, algorithms, and applications. New York: Prentice Hall, 1994.

FERNEDA, Edberto. **Recuperação de informações:** análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. 137 f. Tese (Doutorado em Ciências da Comunicação) - Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo.

FERREIRA, A. B. H (Ed.). **Novo dicionário eletrônico Aurélio versão 5.0**. Rio de Janeiro: Positivo, 2004. 1 CD-ROM. Produzido por Positivo Informática Ltda.

FOX, Christopher. Lexical analysis and stoplists. In: FRAKES, William B, YATES, Ricardo B. **Information retrieval**: data structures and algorithms. Upper Saddle River: Prentice Hall, 1992. p. 1-12.

FRAKES, William B. Introduction to information storage and retrieval system. In: FRAKES, William B, YATES, Ricardo B. **Information retrieval**: data structures and algorithms. Upper Saddle River: Prentice Hall, 1992. p. 1-12.

FRAKES, William B. Stemming algorithms. In: FRAKES, William B, YATES, Ricardo B. **Information retrieval**: data structures and algorithms. Upper Saddle River: Prentice Hall, 1992b. p. 141-167.

FREUND, J. E; SIMON, G. A. **Estatística aplicada:** economia, administração e contabilidade. 9 ed. Porto Alegre: Bookman, 2000.

FUNG, B. C. M; Wang, K; Ester M. **Hierarchical document clustering using frequent itemsets.** Proceedings of the SIAM International Conference on Data Mining, (**SDM 2003**), May 2003, San Francisco, CA.

GIRDWOOD, Ellyn D. **Interface web para submissão de criptogramas a um agrupador hierárquico aglomerativo**. 2006. 60 f. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação). Orientação de William Augusto Rodrigues de Souza. Instituto Metodista Bennett, Rio de Janeiro.

HARMAN, Donna; YATES, Ricardo B; FOX, Edward et al. Inverted files. In: FRAKES, William B, YATES, Ricardo B. **Information retrieval**: data structures and algorithms. Upper Saddle River: Prentice Hall, 1992a. p. 38-53.

HARMAN, Donna. Ranking algorithms. In: FRAKES, William B, YATES, Ricardo B. **Information retrieval**: data structures and algorithms. Upper Saddle River: Prentice Hall, 1992. p. 363-392.

HAYKIN, Simon. **Redes neurais**: princípios e prática. 2 ed. Porto Alegre: Bookman, 2001.

HONKELA, T; KASKI, S; LAGUS, K. et al. **Exploration of Full-Text Databases with Self-Organizing Maps**. 1996. In Proceedings of ICNN'96, IEEE International Conference on Neural Networks, I56-61.

HONKELA[A], T. (1997). **Learning to Understand**: General Aspects of Using Self-Organizing Maps in Natural Language Processing. Computing Anticipatory Systems, D. Dubois (ed.) , American Institute of Physics, Woodbury, New York, pp. 563-576.

HONKELA[B], T. **Self-Organizing Maps in Natural Language Processing**. Ph.D. thesis. Helsinki University of Technology, 1997.

HONKELA[C], Timo; KASKI, Samuel; LAGUS, Krista et al. **WEBSON**: Self-Organizing Maps of Document Collections. In Proceedins of Workshop on Self-Organizing Maps (WSOM-97), Espoo, Finland 1997; 1997:310-5.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data mining**: um guia prático. Rio de Janeiro: Elsevier, 2005.

IEMMA, A. F. **Estatística descritiva**. Piracicaba: Fisigmaro publicações, 1992.

JAIN, A. K; MURTY, M. N; FLYNN, P. J. **Data Clustering:** A Review. ACM Computing Surveys, Vol. 31, No 3, Setember 1999. p. 264-323.

KAHN, David. **The codebreakers**: the story of secret writing. New York: Macmillan Publishing, 1967.

KOHONEN, Teuvo; RITTER, H. **Self-organizing semantic maps**. Biological Cybernetics. 61. Elsevier: Amsterdam, 1989. p. 241-254.

KOHONEN, Teuvo. **Self-organizing maps**. 3 ed. New York: Springer, 2001.

LAMBERT, Jorge de A. **Cifrador simétrico de blocos:** projeto e avaliação. 2004. 353 f. Dissertação (Mestrado em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro.

LIMA, Alexandre A; LIMA, Almir Paz; MARIÑO, Fernando C. C. **Criptoanálise diferencial**. 1999. Relatório Técnico nº RT035/DE9/FEV99 (Mestrado em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro.

MANNING, Christopher D; SCHÜTZE, Hinrich. **Foundations of statistical natural language processing**. Massachusetts: MIT Press, 2003.

MERKL, Dieter. **Exploration of Text Collections with Hierarchical Feature maps**. 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'97). Philadelphia. 1997. p. 186-195.

MENEZES, Alfred J; OORSCHOT, Paul C. Van; VANSTONE, Scott A. **Handbook of applied cryptography**. Boca Raton: CRC Press, 1996.

MEYER, Andréia S. **Comparação de coeficientes de similaridade usados em análise de agrupamento com dados de marcadores moleculares dominantes**. 2002. 106 f. Dissertação (Mestrado em Agronomia) – Universidade de São Paulo, Piracicaba.

MITKOV, Ruslan. **The Oxford Handbook of Computational Linguistics.** New York: Oxford University Press, 2005.

NIST (NATIONAL INSTITUTE OF STANDARD AND TECNOLOGY). **Federal Information Processing Standard, publication 46-3 (FIPS 46-3): Data Encryption Standard (DES)**. Washington D.C., 1999.

NIST (NATIONAL INSTITUTE OF STANDARD AND TECNOLOGY). **Federal Information Processing Standard, publication 197 (FIPS 197): Announcing the advanced encryption standard (AES)**. Washington D.C., 2001.

ODLYZKO, A. M. **The future of integer factorization**. CryptoBytes (The technical newsletter of RSA Laboratories) 1 (no. 2) (1995), pp. 5-12. Disponível: http://www.dtc.umn.edu/~odlyzko/ [capturado 13 dez. 2006].

PBCT/EB (PLANO BÁSICO DE CIÊNCIA E TECNOLOGIA DO EXÉRCITO BRASILEIRO). Disponível: http://www.ime.eb.br [capturado 13 jan. 2007].

RASMUSSEN, Edie. Clustering algorithms. In: FRAKES, William B, YATES, Ricardo B. **Information retrieval**: data structures and algorithms. Upper Saddle River: Prentice Hall, 1992. p. 419-442.

RSA Laboratories. **How large a key should be used in the RSA cryptosystem?.** Disponível: http://www.rsasecurity.com/rsalabs/node.asp?id=2218 [capturado 13 dez. 2006].

SALTON, G; BUCKLEY, C. **Term weighting approaches in automatic text retrieval**. Technical Report, Cornell University, 1987.

SCHEINERMAN, Edward R. **Matemática discreta**: uma introdução. São Paulo: Pioneira Thomson Learning, 2003.

SCHNEIER, Bruce. **Applied cryptography**: protocols, algorithms and source code in C. 2 ed. Massashussets: Addison-Wesley Publishing Company Reading, 1996.

SEBERRY, J; PIEPRZYK, J. **Cryptography:** an introduction to computer security. Upper Saddle River: Prantice Hall, 1989.

SINGH, Simon. **O livro dos códigos**. 3 ed. Rio de Janeiro: Record, 2003.

SOUZA, W. A. R. **Noções matemáticas do algoritmo RSA**. 2005. Trabalho apresentado na disciplina de Criptologia (Mestrado em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro

SOUZA, W. A. R; GOLDSCHMIDT, R. R; Xexéo, J. A. M.; OLIVEIRA, C. M. G. M. de. **Mapa de Kohonen aplicado ao processamento de textos:** casos de estudos em Processamento de Linguagem Natural. 2006. Relatório Técnico (Mestrado em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro (a ser publicado).

STALLINGS, William. **Cryptography and network security:** principles and practice. 2 ed. Upper Saddle River: Prantice Hall, 1998.

SULLIVAN, Terry; NORRIS, Cathleen; PAVUR, Robert. **A Single Dimension of Relevance: Hypothesis Testing the Cluster Hypothesis**. 2006. Disponível: http://www.pantos.org/ts/papers/sdr/ [capturado 11 ago. 2006].

TEKNOMO, Kardi. **Similarity Measurement**. 2006. Disponível: http://people.revoledu.com/kardi/tutorial/Similarity [capturado 20 mar. 2006].

TERADA, Routo. **Segurança de dados:** criptografia em redes de computador. 3 ed. Rio de Janeiro: Record, 2003.

WANNER, Leo. **Introduction to clustering techniques**. 2004. Disponível: www.iula.upf.edu/materials/040701wanner.pdf [capturado 01 mai. 2005].

WARTIK, Steven. Boolean operations. In: FRAKES, William B, YATES, Ricardo B. **Information retrieval**: data structures and algorithms. Upper Saddle River: Prentice Hall, 1992. p. 264-292.

XEXÉO, José A. M; DRAEGER, Vitor H. P; SOUZA, William A. R. **Classificação de documentos com utilização de clusterização hierárquica aglomerativa**. 2006. Relatório Técnico nº RT124/SE-8/MAI06 (Mestrado em Sistemas e Computação) – Instituto Militar de Engenharia, Rio de Janeiro.

YANG, Y; LIU, X. **A re-examination of text categorization methods**. In Proceedings of SIGIR'99, p. 42-49, 1999.

YATES, Ricardo B; NETO, Berthier R. **Modern information retrieval**. New York: Addison Wesley, 1999.

9 APÊNDICES

9.1 APÊNDICE 1: TABELAS COM A ANÁLISE DOS EUROCRYPT DE 1998 A 2003

TAB. 9.1.1: Artigos do Eurocrypt 1998

Conferência	**Total de artigos**	**Artigos sobre criptoanálise**	**Artigos com information retrieval**	**Artigos com o tema proposto**
Eurocrypt 1998	44	8	0	0
Total de textos analisados: 8			**Autores**	
Improved Cryptanalysis of RC5			Alex Biryukov e Eyal Kushilevitz	
Breaking RSA May Not Be Equivalent to Factoring (Extended Abstract)			Dan Boneh e Ramarathnam Venkatesan	
Securing Threshold Cryptosystems against Chosen Ciphertext Attack			Victor Shoup e Rosario Gennaro	
Cryptanalysis of the ANSI X9.52 CBCM Mode			Eli Biham e Lars R. Knudsen	
Differential-Linear Weak Key Classes of IDEA			Philip Hawkes	
Visual Cryptanalysis			Adi Shamir	
Specialized Integer Factorization			Don Coppersmith	
Security of an Identity-Based Cryptosystem and the Related Reductions			Tatsuaki Okamoto e Shigenori Uchiyama	

TAB. 9.1.2: Artigos do Eurocrypt 1999

Conferência	**Total de artigos**	**Artigos sobre criptoanálise**	**Artigos com information retrieval**	**Artigos com o tema proposto**
Eurocrypt 1999	32	6	1	0
Total de textos analisados: 7			**Autores**	
Cryptanalysis of RSA with Private Key d Less than N0.292			Dan Boneh e Glenn Durfee	
An Efficient Threshold Public Key Cryptosystem Secure Against Adaptive Chosen Ciphertext Attack			Ran Canetti e Shafi Goldwasser	
Cryptanalysis of Skipjack Reduced to 31 Rounds Using Impossible Differentials			Eli Biham, Alex Biryukov e Adi Shamir	
Resistance Against General Iterated Attacks			Serge Vaudenay	
Improved Fast Correlation Attacks on Stream Ciphers via Convolutional Codes			Thomas Johansson e Fredrik Jönsson	
Cryptanalysis of an Identification Scheme Based on the Permuted Perceptron Problem			Lars R. Knudsen e Willi Meier	
Computationally Private Information Retrieval with Polylogarithmic Communication			Christian Cachin, Silvio Micali e Markus Stadler	

TAB. 9.1.3: Artigos do Eurocrypt 2000

Conferência	**Total de artigos**	**Artigos sobre criptoanálise**	**Artigos com information retrieval**	**Artigos com o tema proposto**
Eurocrypt 2000	41	11	2	0
Total de textos analisados: 13			**Autores**	
Noisy Polynomial Interpolation and Noisy Chinese Remaindering			Daniel Bleichenbacher e Phong Q. Nguyen	
A Chosen Messages Attack on the ISO/IEC 9796-1 Signature Scheme			François Grieu	
Cryptanalysis of Countermeasures Proposed for Repairing ISO 9796-1			Marc Girault e Jean-François Misarsky	
Security Analysis of the Gennaro-Halevi-Rabin Signature Scheme			Jean-Sébastien Coron e David Naccache	
One-Way Trapdoor Permutations Are Sufficient for Non-trivial Single-Server Private Information Retrieval			Eyal Kushilevitz e Rafail Ostrovsky	
Single Database Private Information Retrieval Implies Oblivious Transfer			Giovanni Di Crescenzo, Tal Malkin, e Rafail Ostrovsky	
Using Hash Functions as a Hedge against Chosen Ciphertext Attack			Victor Shoup	
New Attacks on PKCS#1 v1.5 Encryption			Jean-Sébastien Coron, Marc Joye, David Naccache, e Pascal Paillier	
A NICE Cryptanalysis			Éliane Jaulmes e Antoine Joux	
Efficient Algorithms for Solving Overdefined Systems of Multivariate Polynomial Equations			Nicolas Courtois, Alexander Klimov, Jacques Patarin, e Adi Shamir	
Cryptanalysis of Patarin's 2-Round Public Key System with S Boxes (2R)			Eli Biham	
Improved Fast Correlation Attacks Using Parity-Check Equations of Weight 4 and 5			Anne Canteaut e Michaël Trabbia	
Advanced Slide Attacks			Alex Biryukov e David Wagner	

TAB. 9.1.4: Artigos do Eurocrypt 2001

Conferência	**Total de artigos**	**Artigos sobre criptoanálise**	**Artigos com information retrieval**	**Artigos com o tema proposto**
Eurocrypt 2001	32	4	0	0
Total de textos analisados: 4			**Autores**	
Cryptanalysis of Reduced-Round MISTY			Ulrich Kühn	
The Rectangle Attack - Rectangling the Serpent			Eli Biham, Orr Dunkelman e Nathan Keller	
Key Recovery and Message Attacks on NTRU-Composite			Craig Gentry	
Structural Cryptanalysis of SASAS			Alex Biryukov e Adi Shamir	

TAB. 9.1.5: Artigos do Eurocrypt 2002

Conferência	**Total de artigos**	**Artigos sobre criptoanálise**	**Artigos com information retrieval**	**Artigos com o tema proposto**
Eurocrypt 2002	35	10	0	0
Total de textos analisados: 10			**Autores**	
Fast Correlation Attacks: An Algorithmic Point of View			Philippe Chose, Antoine Joux e Michel Mitton	
BDD-Based Cryptanalysis of Keystream Generators			Matthias Krause	
Linear Cryptanalysis of Bluetooth Stream Cipher			Jovan Golic, Vittorio Bagini e Guglielmo Morgari	
Cryptanalysis of a Pseudorandom Generator Based on Braid Groups			Rosario Gennaro e Daniele Micciancio	
Potential Weaknesses of the Commutator Key Agreement Protocol Based on Braid Groups			Sang Jin Lee e Eonkyung Lee	
Extending the GHS Weil Descent Attack			Steven D. Galbraith, Florian Hess e Nigel P. Smart	
Cryptanalysis of SFLASH			Henri Gilbert e Marine Minier	
Cryptanalysis of the Revised NTRU Signature Scheme			Craig Gentry e Mike Szydlo	
Universal Hash Proofs and a Paradigm for Adaptive Chosen Ciphertext Secure Public-Key Encryption			Ronald Cramer e Victor Shoup	
Degree of Composition of Highly Nonlinear Functions and Applications to Higher Order Differential Cryptanalysis			Anne Canteaut e Marion Videau	

TAB. 9.1.6: Artigos do Eurocrypt 2003

Conferência	**Total de artigos**	**Artigos sobre criptoanálise**	**Artigos com information retrieval**	**Artigos com o tema proposto**
Eurocrypt 2003	39	9	0	0
Total de textos analisados: 9			**Autores**	
Cryptanalysis of the EMD Mode of Operation			Antoine Joux	
On the Optimality of Linear, Differential, and Sequential Distinguishers			Pascal Junod	
A Toolbox for Cryptanalysis: Linear and Affine Equivalence Algorithms			Alex Biryukov, Christophe De Canni`ere, An Braeken e Bart Preneel	
Predicting the Shrinking Generator with Fixed Connections			Patrik Ekdahl, Willi Meier e Thomas Johansson	
Algebraic Attacks on Stream Ciphers with Linear Feedback			Nicolas T. Courtois e Willi Méier	
On the Security of RDSA			Pierre-Alain Fouque e Guillaume Poupard	
Cryptanalysis of the Public-Key Encryption Based on Braid Groups			Eonkyung Lee e Je Hong Park	
A Theoretical Treatment of Related-Key Attacks: RKA-PRPs, RKA-PRFs, and Applications			Mihir Bellare, Tadayoshi Kohno	
The GHS Attack Revisited			Florian Hess	

TAB. 9.1.7: Artigos do Eurocrypt 2004

Conferência	**Total de artigos**	**Artigos sobre criptoanálise**	**Artigos com information retrieval**	**Artigos com o tema proposto**
Eurocrypt 2004	36	2	0	0
Total de textos analisados: 12			**Autores**	
Posititive Results and Techniques for Obfuscation			Benjamin Lynn, Manoj Prabhakaran e Amit Sahai	
Secure Computation of the k^{th} – Ranked Element			Gagan Aggarwal, Nina Mishra e Benny Pinkas	
On the Key-uncertainty of Quantum Ciphers and the Computational Security Of One-Way Quantum Transmission			Ivan Damgård, Thomas Pedersen e Louis Salvail	
The Exact Price for Unconditionally Secure Asymmetric Cryptography			Renato Renner e Stefan Wolf	
On Generating the Initial key in the Bounded-Storage Model			Stefan Dziembowski e Ueli Maurer	
Black–Box Composition Does Not Imply Adaptative Security			Steven Myers	
Chosen-Ciphertext Security from Identity-Based Encryption			Ran Canetti, Shai Halevi e Jonathan Katz	
Efficient Selective-ID Secure Identity-Based Encryption Without Random Oracles			Dan Boneh e Xavier Boyen	
Multi-party Computation with Hybrid Security			Matthias Fitzi, Thomas Holenstein e Jürg Wullschleger	
Algebric Attacks and Decomposition of Boolean Functions			Willi Meier, Enes Pasalic e Claude Carlet	

Finding Small Roots of Bivariate Integer Polynomial Equations Revisited	Jean-Sébastien Coron
Can We Trust Cryptographic Software? Cryptographic Flaws in GNU Privacy Guard v1.2.3	Phong Q. Nguyen

TAB. 9.1.8: Artigos do Eurocrypt 2006

Conferência	**Total de artigos**	**Artigos sobre criptoanálise**	**Artigos com information retrieval**	**Artigos com o tema proposto**
Eurocrypt 2006	35	0	0	0
Total de textos analisados: 8			**Autores**	
Security Analysis of the Strong Diffie-Hellman Problem			Jung Hee Cheon	
Hiding Secret Points Amidst Chaff			Ee-Chien Chang e Qiming Li	
QUAD: A Practical Stream Cipher with Provable Security			Côme Berbain, Henri Gilbert e Jacques Patarin	
Efficient Computation of Algebraic Immunity for Algebraic and Fast Algebraic Attacks			Freredik Armknecht, Claude Carlet, Philippe Gaborit, Simon Künzli, Willi Meier e Olivier Ruatta	
Herding Hash Functions and the Nostradamus Attack			John Kelsey e Tadayoshi Kohno	
Learning a Parallelepiped: Cryptanalysis of GGH and NTRU Signatures			Phong Q. Nguyen e Oded Regev	
Language Modeling and Encryption on Packet Switched Networks			Kevin S. McCurley	
Simplified Threshold RSA with Adaptive and Proactive Security			Jesús F. Almansa, Ivan Damgård e Jesper Burus Nielsen	

9.2 APÊNDICE 2: MAIORES VALORES DE DISSIMILARIDADE

9.2.1 VALORES PARA O ALGORITMO DES

Tamanho dos textos (bytes)	Euclidiana	Manhattan	Canberra	Bray-Curtis
10240	60,46	2560	2466	1
8192	55,75	2048	1990	1
6144	42,66	1536	1498	1
4096	35,16	1024	1012	1
2048	26,46	512	510	1
1024	20,2	256	256	1
512	26,46	128	128	1
256	8,94	64	64	1
192	7,48	48	48	1
128	6	32	32	1
64	4,47	16	16	1

9.2.2 VALORES PARA O ALGORITMO AES COM CHAVE DE 128 BITS

Tamanho dos textos (bytes)	Euclidiana	Manhattan	Canberra	Bray-Curtis
10240	36,44	1280	1280	1
8192	32,56	1024	1024	1
6144	28,43	768	768	1
4096	23,41	512	512	1
3072	20,4	384	384	1
2560	18,76	320	320	1
2048	17,09	256	256	1
1536	14,97	192	192	1
1024	12,65	128	128	1

9.2.3 VALORES PARA O ALGORITMO AES COM CHAVE DE 192 BITS

Tamanho dos textos (bytes)	Euclidiana	Manhattan	Canberra	Bray-Curtis
10240	29,5	854	854	1
8192	26,46	684	684	1
6144	22,99	512	512	1
4096	18,92	342	342	1
3072	16,49	256	256	1
2560	15,17	214	214	1
2048	13,71	172	172	1
1536	12	128	128	1
1024	10,1	86	86	1

9.2.4 VALORES PARA O ALGORITMO AES COM CHAVE DE 256 BITS

Tamanho dos textos (bytes)	Euclidiana	Manhattan	Canberra	Bray-Curtis
10240	25,53	640	640	1
8192	22,89	512	512	1
6144	19,9	384	384	1
4096	16,37	256	256	1
3072	14,28	192	192	1
2560	13,11	160	160	1
2048	11,83	128	128	1
1536	10,39	96	96	1
1024	8,72	64	64	1

9.2.5 VALORES PARA O ALGORITMO RSA COM CHAVE DE 64 BITS

Tamanho dos textos (bytes)	Euclidiana	Manhattan	Canberra	Bray-Curtis
10240	70,84	2926	2646	1
8192	60,38	2342	2196	1
6144	54,37	1756	1658	1
4096	49,27	1172	1114	1
2048	26,57	586	580	1
1024	18,92	294	292	1
512	12,96	148	148	1
256	8,83	74	74	1
192	8	56	56	1
128	6,48	38	38	1
64	4,47	20	20	1

9.2.6 VALORES PARA O ALGORITMO RSA COM CHAVE DE 128 BITS

Tamanho dos textos (bytes)	Euclidiana	Manhattan	Canberra	Bray-Curtis
10240	38,29	1366	1366	1
8192	34,09	1094	1094	1
6144	28,91	820	820	1
4096	25,92	548	548	1
2048	17,15	274	274	1
1024	12,08	138	138	1
512	8,37	70	70	1
256	6	36	36	1
192	5,48	26	26	1
128	4,24	18	18	1
64	3,16	10	10	1

9.2.7 VALORES PARA O ALGORITMO RSA COM CHAVE DE 256 BITS

Tamanho dos textos (bytes)	Euclidiana	Manhattan	Canberra	Bray-Curtis
10240	26,04	662	662	1
8192	23,11	530	530	1
6144	19,95	398	398	1
4096	16,67	266	266	1
2048	11,75	134	134	1
1024	8,25	68	68	1
512	5,83	34	34	1
256	4,24	18	18	1
192	3,74	14	14	1
128	3,16	10	10	1
64	2,45	6	6	1

9.2.8 VALORES PARA O ALGORITMO RSA COM CHAVE DE 512 BITS

Tamanho dos textos (bytes)	Euclidiana	Manhattan	Canberra	Bray-Curtis
10240	18,06	326	326	1
8192	16,19	262	262	1
6144	14	196	196	1
4096	11,49	132	132	1
2048	8,12	66	66	1
1024	5,83	34	34	1
512	4,24	18	18	1
256	3,16	10	10	1
192	2,83	8	8	1
128	2,45	6	6	1
64	2	4	4	1

9.2.9 VALORES PARA O ALGORITMO RSA COM CHAVE DE 1024 BITS

Tamanho dos textos (bytes)	Euclidiana	Manhattan	Canberra	Bray-Curtis
10240	12,73	162	162	1
8192	11,4	130	130	1
6144	9,9	98	98	1
4096	8,12	66	66	1
2048	5,83	34	34	1
1024	4,24	18	18	1
512	3,16	10	10	1
256	2,45	6	6	1
192	2	4	4	1
128	2	4	4	1
64	1,41	2	2	1

9.2.10 VALORES PARA OS DEMAIS EXPERIMENTOS

Experimento	Euclidiana	Manhattan	Canberra	Bray-Curtis
DES (variado)	29,97	768	761	1
Cifra 64 bits (modo ECB)	310,68	2560	2366	1
Cifra 128 bits (modo ECB)	144,45	1280	1256	1
Cifra 128 bits (modo CBC)	35,78	1280	1280	1

9.3 APÊNDICE 3: PRIMEIRO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO DES

9.3.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	total
10240	1	1	705,2	0,86	0,06	706,08	1	1	700,1	0,88	0,14	701,14
8192	1	1	554,8	0,86	0,09	555,80	1	1	550,6	0,94	0,08	551,61
6144	1	1	426,4	0,86	0,02	427,28	1	1	420,9	0,84	0,02	421,73
4096	1	1	278,5	0,89	0,02	279,42	1	1	275,0	0,92	0,02	275,98
2048	1	1	142,4	0,88	0,03	143,28	1	1	139,7	0,81	0,02	140,48
1024	1	1	73,2	0,78	0,03	74,00	1	1	70,8	0,86	0,02	71,69
512	1	1	39,1	0,91	0,02	39,99	1	1	37,5	0,78	0,02	38,28
256	1	0,8	21,1	1,2	0,02	22,28	1	0,8	20,0	1,19	0,02	21,25
192	1	0,73	15,7	1,31	0,02	17,02	1	0,73	15,8	1,34	0,02	17,11
128	1	0,4	11,5	3,44	0,03	14,96	1	0,4	11,5	3,47	0,03	15,00
64	1	0,6	8,1	2,08	0,17	10,39	1	0,6	7,4	1,92	0,02	9,31

	Jaccard						Overlap					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	Total
10240	1	1	699,9	0,88	0,03	700,84	1	1	700,7	0,86	0,03	701,63
8192	1	1	548,6	0,89	0,19	549,67	1	1	550,9	0,86	0,06	551,84
6144	1	1	424,1	0,78	0,05	424,91	1	1	424,4	0,8	0,02	425,17
4096	1	1	277,8	0,8	0,02	278,66	1	1	277,5	0,8	0,02	278,30
2048	1	1	139,4	0,83	0,05	140,23	1	1	141,7	0,81	0,02	142,53
1024	1	1	70,4	0,78	0,02	71,17	1	1	71,6	0,73	0,02	72,34
512	1	1	37,3	0,77	0,03	38,09	1	1	37,8	0,78	0,03	38,64
256	1	0,8	20,0	1,34	0,03	21,42	1	0,8	20,3	1,36	0,02	21,72
192	1	0,73	15,8	1,36	0,02	17,20	1	0,73	16,1	1,25	0,03	17,38
128	1	0,4	11,5	3,53	0,02	15,06	1	0,4	11,7	3,5	0,02	15,17
64	1	0,6	8,4	2,11	0,05	10,52	1	0,6	7,6	2,09	0,03	9,69

	Simple-matching					
Tamanho dos textos (bytes)			Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total
10240	1	1	726,7	0,92	0,05	727,63
8192	1	1	550,2	0,88	0,06	551,16
6144	1	1	431,0	0,81	0,03	431,84
4096	1	1	278,0	0,84	0,13	278,92
2048	1	1	143,0	0,84	0,03	143,84
1024	1	1	72,1	1,11	0,03	73,22
512	1	1	37,8	0,78	0,03	38,56
256	1	0,8	20,1	1,41	0,02	21,55
192	1	0,73	15,5	1,25	0,02	16,81
128	1	0,4	11,5	3,49	0,03	15,05
64	1	0,6	7,1	2,08	0,03	9,23

9.3.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	total
10240	0,09	1	4.680,7	6,19	0,02	4.686,89	1	1	4.098,1	6,41	0,06	4.104,53
8192	0,05	1	3.791,6	6,11	0,08	3.797,81	1	1	3.229,9	6,58	0,08	3.236,52
6144	0,02	1	2.794,3	6,16	0,03	2.800,45	1	1	2.417,6	6,27	0,09	2.423,99
4096	0,05	1	1.764,1	6,13	0,03	1.770,30	1	1	1.526,5	6,39	0,03	1.532,88
2048	0,02	1	732,6	6,81	0,02	739,39	1	0,43	613,7	6,34	0,05	620,11
1024	0,02	1	311,0	6,81	0,02	317,81	1	0,13	252,1	6,17	0,03	258,30
512	0,05	1	151,8	6,78	0,02	158,59	1	0,1	122,4	6,39	0,05	128,83
256	0,15	1	70,8	6,52	0,03	77,36	1	0,13	58,4	6,45	0,05	64,89
192	0,09	1	52,7	6,52	0,02	59,24	1	0,13	42,9	6,45	0,06	49,39
128	0,25	0,93	34,8	6,38	0,03	41,24	1	0,1	28,1	6,42	0,05	34,58
64	0,18	0,9	17,4	6,92	0,05	24,34	1	0,13	13,7	6,2	0,02	19,91

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	total
10240	1	1	2.319,3	5,953	0,047	2.325,31	1	1	4.057,1	1,00	0,031	4.058,11
8192	1	1	1.873,2	6,437	0,046	1.879,72	1	1	3.221,5	0,922	0,078	3.222,48
6144	1	1	1.339,8	6,469	0,062	1.346,31	1	1	2.386,1	1,078	0,063	2.387,27
4096	1	1	703,5	6,203	0,047	709,75	1	0,67	1.527,1	0,906	0,015	1.528,00
2048	0,97	1	245,3	6,89	0,016	252,16	1	0,2	602,9	0,86	0,032	603,78
1024	0,9	0,47	112,1	7,625	0,046	119,81	1	0,1	254,2	4,00	0,015	258,25
512	0,77	0,17	56,3	7,266	0,063	63,64	1	0,1	122,0	5,625	0,047	127,72
256	0,87	0,13	28,8	7,266	0,046	36,12	1	0,13	57,8	5,968	0,031	63,81
192	0,8	0,2	21,2	7,531	0,047	28,80	1	0,13	42,9	5,484	0,032	48,45
128	1	0,13	14,8	9,407	0,078	24,31	1	0,1	27,5	5,641	0,016	33,13
64	1	0,13	8,1	8,985	0,015	17,06	1	0,13	13,5	5,969	0,047	19,49

9.3.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	total
10240	1	1	705,16	0,69	0,02	705,86	1	1	700,13	0,69	0,19	701,00
8192	1	1	554,84	0,69	0,05	555,58	1	1	550,59	0,69	0,08	551,36
6144	1	1	426,41	0,69	0,02	427,11	1	1	420,88	0,69	0,02	421,58
4096	1	1	278,52	0,70	0,02	279,23	1	1	275,05	0,70	0,02	275,76
2048	1	0,5	142,38	0,78	0,02	143,17	1	0,5	139,66	0,78	0,03	140,47
1024	1	0,17	73,19	1,47	0,02	74,67	1	0,17	70,81	1,48	0,03	72,33
512	1	0,10	39,06	2,75	0,05	41,86	1	0,10	37,49	2,70	0,02	40,20
256	1	0,13	21,06	3,94	0,02	25,01	1	0,13	20,05	4,23	0,03	24,31
192	1	0,17	15,69	3,74	0,03	19,45	1	0,17	15,75	3,73	0,05	19,53
128	1	0,10	11,49	5,06	0,03	16,58	1	0,10	11,50	5,09	0,02	16,61
64	1	0,13	8,14	3,48	0,05	11,67	1	0,13	7,38	3,49	0,06	10,92

Tamanho dos textos (bytes)	Jaccard						Overlap					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	Total
10240	1	1	699,94	0,70	0,03	700,67	1	1	700,73	0,69	0,03	701,45
8192	1	1	548,59	0,70	0,19	549,48	1	1	550,92	0,70	0,08	551,71
6144	1	1	424,08	0,70	0,02	424,80	1	1	424,36	0,67	0,02	425,05
4096	1	1	277,84	0,70	0,03	278,58	1	1	277,49	0,69	0,03	278,20
2048	1	0,5	139,36	0,78	0,03	140,17	1	0,5	141,70	0,78	0,05	142,53
1024	1	0,17	70,38	1,47	0,05	71,89	1	0,17	71,59	1,30	0,02	72,91
512	1	0,10	37,30	2,74	0,02	40,05	1	0,10	37,83	2,66	0,03	40,52
256	1	0,13	20,05	3,91	0,02	23,97	1	0,13	20,34	3,92	0,05	24,31
192	1	0,17	15,83	3,77	0,05	19,64	1	0,17	16,09	3,74	0,03	19,86
128	1	0,10	11,52	5,08	0,03	16,63	1	0,10	11,66	5,13	0,06	16,84
64	1	0,13	8,36	3,41	0,03	11,80	1	0,13	7,56	3,44	0,03	11,03

Tamanho dos textos (bytes)	Simple-matching					
			Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total
10240	1	1	726,66	0,70	0,13	727,48
8192	1	1	550,22	0,69	0,05	550,95
6144	1	1	431,00	0,70	0,05	431,75
4096	1	1	277,95	0,69	0,08	278,72
2048	1	0,4	142,97	0,77	0,02	143,75
1024	1	0,17	72,08	1,39	0,03	73,50
512	1	0,10	37,75	2,84	0,02	40,61
256	1	0,13	20,13	3,94	0,02	24,08
192	1	0,17	15,55	3,75	0,03	19,33
128	1	0,10	11,53	5,05	0,03	16,61
64	1	0,13	7,13	3,41	0,02	10,55

9.3.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	total
10240	0,02	1	4.680,69	5,95	0,06	4.686,70	1	1	4.098,06	6,00	0,05	4.104,11
8192	0,02	1	3.791,63	5,99	0,09	3.797,70	1	1	3.229,86	6,09	0,05	3.236,00
6144	0,02	1	2.794,27	5,97	0,06	2.800,30	1	1	2.417,63	6,06	0,08	2.423,77
4096	0,02	1	1.764,14	6,05	0,03	1.770,22	1	1	1.526,45	6,05	0,03	1.532,53
2048	0,02	1	732,56	6,36	0,05	738,97	1	1	613,72	5,08	0,03	618,83
1024	0,02	1	310,99	5,67	0,05	316,70	1	1	252,09	4,78	0,05	256,92
512	0,02	1	151,80	5,49	0,05	157,33	1	1	122,39	5,27	0,02	127,67
256	0,02	1	70,81	5,50	0,03	76,34	1	0,8	58,39	4,97	0,03	63,39
192	0,02	1	52,70	5,66	0,05	58,41	1	0,73	42,88	4,66	0,03	47,56
128	0,02	1	34,83	5,55	0,03	40,41	1	0,4	28,11	5,56	0,05	33,72
64	0,02	1	17,38	5,91	0,05	23,33	1	0,6	13,69	5,23	0,05	18,97

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	total
10240	0,02	1	2.319,31	5,83	0,03	2.325,17	1	1	4.057,08	0,73	0,03	4.057,84
8192	0,02	1	1.873,23	6,16	0,08	1.879,47	1	1	3.221,48	0,77	0,05	3.222,30
6144	0,02	1	1.339,78	6,13	0,08	1.345,98	1	1	2.386,13	0,84	0,08	2.387,05
4096	0,02	1	703,50	6,11	0,03	709,64	1	1	1.527,08	1,33	0,03	1.528,44
2048	0,02	1	245,25	6,61	0,06	251,92	1	1	602,89	2,75	0,03	605,67
1024	0,02	1	112,14	5,55	0,02	117,70	1	1	254,23	3,86	0,06	258,16
512	0,02	1	56,31	5,33	0,05	61,69	1	1	122,05	4,58	0,02	126,64
256	0,02	1	28,81	5,89	0,03	34,74	1	0,8	57,81	4,41	0,03	62,25
192	0,02	1	21,22	6,19	0,05	27,45	1	0,733	42,94	5,08	0,05	48,06
128	0,608	0,4	14,83	8,03	0,03	22,89	1	0,4	27,47	5,06	0,06	32,59
64	0,412	0,6	8,06	7,50	0,06	15,63	1	0,6	13,47	4,80	0,02	18,28

9.3.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Indexação	Clust.	Análise	total			Indexação	Clust.	Análise	total
10240	1	1	705,16	0,59	0,03	705,78	1	1	700,13	0,61	0,03	700,77
8192	1	1	554,84	0,63	0,08	555,55	1	1	550,59	0,61	0,06	551,27
6144	1	1	426,41	0,61	0,02	427,03	1	1	420,88	0,59	0,03	421,50
4096	1	1	278,52	0,59	0,05	279,16	1	1	275,05	0,59	0,03	275,67
2048	1	1	142,38	0,61	0,08	143,06	1	1	139,66	0,61	0,05	140,31
1024	1	1	73,19	0,61	0,05	73,84	1	1	70,81	0,59	0,02	71,42
512	1	1	39,06	0,58	0,02	39,66	1	1	37,49	0,59	0,02	38,09
256	1	0,8	21,06	0,91	0,03	22,00	1	0,8	20,05	0,92	0,03	21,00
192	1	0,5	15,69	1,02	0,02	16,72	1	0,5	15,75	1,02	0,02	16,78
128	1	0,4	11,49	3,30	0,03	14,81	1	0,4	11,50	3,31	0,05	14,86
64	1	0,6	8,14	1,86	0,05	10,05	1	0,6	7,38	1,89	0,06	9,33

Tamanho dos textos (bytes)	Jaccard						Overlap					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Indexação	Clust.	Análise	total			Indexação	Clust.	Análise	total
10240	1	1	699,94	0,59	0,03	700,56	1	1	700,73	0,63	0,05	701,41
8192	1	1	548,59	0,59	0,06	549,25	1	1	550,92	0,61	0,05	551,58
6144	1	1	424,08	0,59	0,03	424,70	1	1	424,36	0,61	0,02	424,98
4096	1	1	277,84	0,59	0,03	278,47	1	1	277,49	0,61	0,02	278,11
2048	1	1	139,36	0,59	0,03	139,98	1	1	141,70	0,61	0,03	142,35
1024	1	1	70,38	0,59	0,05	71,02	1	1	71,59	0,59	0,02	72,20
512	1	0,57	37,30	0,59	0,02	37,91	1	1	37,83	0,58	0,02	38,42
256	1	0,33	20,05	0,92	0,02	20,99	1	0,8	20,34	0,91	0,02	21,27
192	1	0,37	15,83	1,03	0,02	16,87	1	0,5	16,09	1,03	0,03	17,16
128	1	0,4	11,52	3,30	0,05	14,86	1	0,4	11,66	3,25	0,05	14,95
64	1	0,6	8,36	1,86	0,05	10,27	1	0,6	7,56	1,83	0,03	9,42

Tamanho dos textos (bytes)	Simple-matching					
	P	R	Tempo (segundos)			
			Indexação	Clust.	Análise	total
10240	1	1	726,66	0,59	0,06	727,31
8192	1	1	550,22	0,63	0,05	550,89
6144	1	1	431,00	0,58	0,05	431,63
4096	1	1	277,95	0,61	0,05	278,61
2048	1	1	142,97	0,59	0,05	143,61
1024	1	1	72,08	0,58	0,05	72,70
512	1	1	37,75	0,58	0,03	38,36
256	1	0,8	20,13	0,91	0,02	21,05
192	1	0,73	15,55	1,02	0,05	16,61
128	1	0,4	11,53	3,25	0,05	14,83
64	1	0,6	7,13	1,84	0,03	9,00

9.3.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	total
10240	0,02	1	4.680,69	3,88	0,02	4.684,58	1	1	4.098,06	6,02	0,05	4.104,12
8192	0,02	1	3.791,63	3,52	0,08	3.795,22	1	1	3.229,86	6,02	0,33	3.236,21
6144	0,02	1	2.794,27	2,92	0,05	2.797,24	1	1	2.417,63	6,02	0,09	2.423,74
4096	0,02	1	1.764,14	2,34	0,03	1.766,52	1	1	1.526,45	6,00	0,03	1.532,49
2048	0,02	1	732,56	2,41	0,05	735,02	1	1	613,72	5,05	0,05	618,81
1024	0,02	1	310,99	2,02	0,03	313,03	1	1	252,09	4,69	0,03	256,81
512	0,02	1	151,80	1,78	0,02	153,59	1	1	122,39	5,30	0,06	127,75
256	0,02	1	70,81	1,30	0,05	72,16	1	0,8	58,39	4,94	0,05	63,38
192	0,02	1	52,70	1,19	0,02	53,91	1	0,73	42,88	4,63	0,05	47,55
128	0,02	1	34,83	1,41	0,05	36,28	1	0,4	28,11	5,52	0,03	33,66
64	0,02	1	17,38	1,14	0,05	18,56	1	0,6	13,69	5,19	0,05	18,92

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Indexação	Clust.	Análise	total	P	R	Indexação	Clust.	Análise	total
10240	0,2	1	2.319,31	2,86	0,05	2.322,22	1	1	4.057,08	0,74	0,03	4.057,84
8192	0,2	1	1.873,23	4,02	0,06	1.877,31	1	1	3.221,48	0,73	0,06	3.222,28
6144	0,2	1	1.339,78	2,58	0,20	1.342,56	1	1	2.386,13	0,73	0,05	2.386,91
4096	0,2	1	703,50	3,05	0,05	706,59	1	1	1.527,08	0,72	0,03	1.527,83
2048	0,2	1	245,25	2,92	0,11	248,28	1	0,93	602,89	0,72	0,03	603,64
1024	0,2	1	112,14	3,28	0,06	115,49	1	0,2	254,23	3,67	0,14	258,05
512	0,2	1	56,31	4,92	0,03	61,27	1	0,17	122,05	5,28	0,06	127,39
256	0,78	0,8	28,81	5,83	0,05	34,69	1	0,2	57,81	6,27	0,06	64,14
192	0,74	0,4	21,22	5,94	0,05	27,20	1	0,23	42,94	5,48	0,02	48,44
128	1	0,3	14,83	7,88	0,05	22,75	1	0,33	27,47	6,11	0,05	33,63
64	1	0,2	8,06	7,63	0,02	15,70	1	0,5	13,47	6,52	0,03	20,02

9.3.7 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE COM A FORMAÇÃO DE 50 GRUPOS COMO CRITÉRIO DE PARADA

9.3.7.1 MÉTODO SINGLE-LINK

	Euclidiana		Manhattan		Canberra		Bray-Curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
10240	0,09	1	1	1	1	1	1	1
8192	0,05	1	1	1	1	1	1	1
6144	1	1	1	1	1	1	1	1
4096	0,05	1	1	1	1	1	1	1
2048	0,05	1	0,216	1	0,967	1	0,22	1
1024	0,05	1	0,118	1	0,315	0,9	0,12	1
512	0,05	1	0,082	1	0,089	0,767	0,08	1
256	0,15	1	0,072	1	0,067	0,867	0,07	1
192	0,09	1	0,07	1	0,064	0,8	0,07	1
128	0,14	0,93	0,066	1	0,068	1	0,07	1
64	0,1	0,9	0,073	1	0,073	1	0,07	1

9.3.7.2 MÉTODO COMPLETE-LINK

	Euclidiana		Manhattan		Canberra		Bray-Curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
10240	0,053	1	1	1	1	1	1	1
8192	0,053	1	1	1	1	1	1	1
6144	0,053	1	1	1	1	1	1	1
4096	0,053	1	1	1	0,053	1	1	1
2048	0,053	1	1	1	0,053	1	1	1
1024	0,053	1	1	1	0,053	1	1	1
512	0,069	1	1	1	0,053	1	1	1
256	0,14	1	0,16	1	0,058	1	0,16	1
192	0,1	1	0,14	1	0,088	1	0,14	1
128	0,075	1	0,077	1	0,055	1	0,077	1
64	0,095	1	0,098	1	0,053	1	0,098	1

9.3.7.3 MÉTODO GROUP AVERAGE-LINK

	Euclidiana		Manhattan		Canberra		Bray-Curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
10240	0,053	1	1	1	1	1	1	1
8192	0,053	1	1	1	1	1	1	1
6144	1	1	1	1	1	1	1	1
4096	0,053	1	1	1	1	1	1	1
2048	0,053	1	1	1	0,053	1	1	1
1024	0,053	1	1	1	0,053	1	1	1
512	0,053	1	1	1	0,053	1	1	1
256	0,15	1	0,16	1	0,053	1	0,16	1
192	0,085	1	0,14	1	0,057	1	0,14	1
128	0,347	1	0,076	1	0,055	1	0,077	1
64	0,085	1	0,098	1	0,053	1	0,098	1

9.4 APÊNDICE 4: PRIMEIRO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO AES, COM CHAVES 128 BITS

9.4.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)	P	R	Tempo				P	R	Tempo			
			Index	Clus	Analis	Total			Index	Clus	Analis	total
10240	1	1	376,10	0,89	0,01	377,01	1	1	378,24	0,88	0,03	379,14
8192	1	1	296,01	0,78	0,01	296,80	1	1	294,23	0,80	0,01	295,04
6144	1	1	222,22	0,95	0,01	223,19	1	1	218,92	0,92	0,02	219,86
4096	1	0,97	149,66	0,94	0,01	150,61	1	0,97	146,54	0,88	0,01	147,43
3072	1	0,87	110,21	0,94	0,01	111,16	1	0,87	107,82	1,00	0,01	108,83
2560	1	0,5	93,76	1,25	0,02	95,03	1	0,5	91,72	1,16	0,01	92,89
2048	1	0,33	75,74	2,66	0,03	78,43	1	0,33	72,72	2,88	0,02	75,61
1536	1	0,2	57,49	4,13	0,03	61,65	1	0,2	54,58	5,32	0,02	59,91
1024	1	0,17	39,49	6,78	0,16	46,43	1	0,17	37,88	7,07	0,04	44,99

	Jaccard						Overlap					
Tamanho dos textos (bytes)	P	R	Tempo				P	R	Tempo			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	1	373,62	1,19	0,01	374,82	1	1	375,31	0,84	0,01	376,17
8192	1	1	292,81	0,81	0,01	293,63	1	1	295,38	0,77	0,01	296,16
6144	1	1	219,91	0,95	0,03	220,90	1	1	219,98	0,88	0,01	220,87
4096	1	0,97	146,50	0,92	0,07	147,49	1	0,97	146,92	0,98	0,01	147,92
3072	1	0,87	108,69	1,13	0,02	109,83	1	0,87	100,92	0,92	0,03	101,87
2560	1	0,5	91,05	1,17	0,02	92,24	1	0,5	92,28	1,08	0,01	93,37
2048	1	0,33	72,47	3,08	0,02	75,57	1	0,33	72,49	2,80	0,02	75,31
1536	1	0,2	54,97	5,04	0,03	60,05	1	0,2	55,14	4,98	0,03	60,16
1024	1	0,17	37,89	6,93	0,02	44,84	1	0,17	38,18	7,65	0,02	45,86

	Simple-matching					
Tamanho dos textos (bytes)	P	R	Tempo			
			Index	Clus	Analis	total
10240	1	1	377,71	1,03	0,01	378,75
8192	1	1	299,88	0,81	0,03	300,72
6144	1	1	224,01	0,77	0,01	224,79
4096	1	0,97	149,25	0,86	0,03	150,13
3072	1	0,87	109,23	1,00	0,01	110,25
2560	1	0,5	91,39	1,16	0,02	92,56
2048	1	0,33	74,28	2,36	0,02	76,66
1536	1	0,2	55,80	4,85	0,03	60,69
1024	1	0,17	38,22	6,18	0,03	44,43

9.4.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,053	1	2.348,12	6,55	0,11	2.354,78	1	0,13	2.058,98	6,30	0,02	2.065,29
8192	0,053	1	1.796,53	6,78	0,02	1.803,33	1	0,2	1.585,85	6,14	0,04	1.592,02
6144	0,053	1	1.277,54	6,47	0,04	1.284,05	1	0,13	1.131,24	6,22	0,04	1.137,49
4096	0,053	1	721,86	6,88	0,02	728,75	1	0,1	626,97	6,31	0,04	633,32
3072	0,053	1	522,49	8,\016	0,02	522,51	1	0,1	438,50	6,31	0,04	444,85
2560	0,053	1	408,52	6,36	0,02	414,90	1	0,07	341,45	6,42	0,04	347,91
2048	0,053	1	316,33	7,00	0,02	323,35	1	0,07	263,09	6,89	0,04	270,02
1536	0,053	1	239,25	7,48	0,04	246,77	1	0,07	198,81	6,96	0,02	205,79
1024	0,053	1	155,34	9,45	0,04	164,83	1	0,07	127,78	7,76	0,02	135,57

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,67	0,5	1.113,11	6,49	0,02	1.119,62	1	0,07	2.054,45	3,97	0,04	2.058,46
8192	0,48	0,53	769,50	6,44	0,02	775,95	1	0,07	1.575,01	4,48	0,04	1.579,53
6144	0,94	0,4	482,79	7,27	0,04	490,09	1	0,07	1.137,05	5,05	0,05	1.142,15
4096	0,68	0,3	257,53	7,23	0,02	264,78	1	0,07	627,79	6,00	0,04	633,83
[redacted]									437,72	6,13	0,04	443,88
									337,38	6,19	0,02	343,59
									261,46	6,45	0,02	267,93
									194,77	7,08	0,04	201,89
									126,68	6,15	0,04	132,87

9.4.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,13	376,10	1,63	0,02	377,75	1	0,13	378,24	1,64	0,01	379,89
8192	1	0,17	296,01	1,50	0,01	297,52	1	0,17	294,23	1,49	0,02	295,73
6144	1	0,1	222,22	2,41	0,02	224,65	1	0,1	218,92	2,53	0,02	221,47
4096	1	0,1	149,66	2,73	0,03	152,42	1	0,1	146,54	2,58	0,06	149,17
3072	1	0,1	110,21	3,30	0,02	113,52	1	0,1	107,82	3,23	0,03	111,08
2560	1	0,07	93,76	3,58	0,02	97,36	1	0,07	91,72	3,58	0,03	95,33
2048	1	0,07	75,74	4,22	0,02	79,98	1	0,07	72,72	4,20	0,02	76,94

9.4.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,02	1	2.348,12	5,63	0,02	2.353,76	1	1	2.058,98	3,69	0,04	2.062,70
8192	0,02	1	1.796,53	6,17	0,05	1.802,74	1	1	1.585,85	4,31	0,02	1.590,18
6144	0,02	1	1.277,54	5,58	0,02	1.283,14	1	1	1.131,24	4,52	0,04	1.135,79
4096	0,02	1	721,86	5,95	0,04	727,85	1	0,97	626,97	4,00	0,02	630,99
3072	0,02	1	522,49	5,11	0,04	527,64	1	0,87	438,50	4,75	0,02	443,28
2560	0,02	1	408,52	5,61	0,04	414,16	1	0,5	341,45	4,92	0,02	346,39
2048	0,02	1	316,33	6,03	0,04	322,40	1	0,33	263,09	5,48	0,02	268,60
1536	0,02	1	239,25	6,24	0,02	245,51	1	0,2	198,81	5,66	0,04	204,50
1024	0,02	1	155,34	6,56	0,02	161,92	1	0,17	127,78	5,45	0,04	133,27

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,02	1	1.113,11	5,28	0,04	1.118,43	1	0,2	2.054,45	3,95	0,02	2.058,42
8192	0,02	1	769,50	5,42	0,04	774,96	1	0,07	1.575,01	4,25	0,02	1.579,28
6144	0,02	1	482,79	5,77	0,02	488,57	1	0,1	1.137,05	4,14	0,02	1.141,21
4096	0,02	1	257,53	6,05	0,04	263,62	1	0,07	627,79	3,55	0,02	631,36
3072	0,02	1	181,50	5,11	0,04	186,65	1	0,1	437,72	4,89	0,02	442,63
2560	0,02	1	146,74	6,20	0,02	152,97	1	0,1	337,38	5,03	0,04	342,45
2048	0,02	1	115,36	6,36	0,02	121,74	1	0,07	261,46	5,45	0,04	266,95
1536	0,02	1	87,00	7,03	0,04	94,06	1	0,2	194,77	5,71	0,02	200,50
1024	0,02	1	56,83	6,96	0,04	63,83	1	0,17	126,68	5,52	0,02	132,22

9.4.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	1	376,10	0,59	0,01	376,71	1	1	378,24	0,61	0,01	378,86
8192	1	1	296,01	0,59	0,03	296,63	1	1	294,23	0,59	0,01	294,84
6144	1	1	222,22	0,61	0,01	222,85	1	1	218,92	0,61	0,01	219,54
4096	1	0,97	149,66	0,61	0,01	150,28	1	0,97	146,54	0,61	0,05	147,19
3072	1	0,87	110,21	0,66	0,01	110,88	1	0,87	107,82	0,64	0,02	108,47
2560	1	0,5	93,76	0,77	0,01	94,54	1	0,5	91,72	0,78	0,01	92,51
2048	1	0,33	75,74	1,91	0,02	77,66	1	0,33	72,72	1,91	0,03	74,66
1536	1	0,2	57,49	3,72	0,02	61,23	1	0,2	54,58	3,70	0,02	58,30
1024	1	0,17	39,49	4,42	0,02	43,93	1	0,17	37,88	4,35	0,04	42,28

	Jaccard						Overlap					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	1	373,62	0,61	0,01	374,24	1	1	375,31	0,61	0,01	375,94
8192	1	1	292,81	0,59	0,03	293,43	1	1	295,38	0,59	0,01	295,99
6144	1	1	219,91	0,61	0,01	220,54	1	1	219,98	0,59	0,01	220,59
4096	1	0,97	146,50	0,59	0,01	147,11	1	0,97	146,92	0,59	0,01	147,53
3072	1	0,87	108,69	0,66	0,02	109,36	1	0,87	100,92	0,64	0,01	101,58
2560	1	0,5	91,05	0,78	0,03	91,86	1	0,5	92,28	0,75	0,01	93,04
2048	1	0,33	72,47	1,91	0,02	74,40	1	0,33	72,49	1,89	0,02	74,40
1536	1	0,2	54,97	3,72	0,02	58,71	1	0,2	55,14	3,66	0,02	58,82
1024	1	0,17	37,89	4,40	0,02	42,30	1	0,17	38,18	4,36	0,02	42,56

	Simple-matching					
Tamanho dos textos (bytes)			Tempo			
	P	R	Index	Clus	Analis	total
10240	1	1	377,71	0,59	0,01	378,32
8192	1	1	299,88	0,58	0,01	300,47
6144	1	1	224,01	0,59	0,03	224,64
4096	1	0,97	149,25	0,58	0,01	149,84
3072	1	0,87	109,23	0,63	0,01	109,87
2560	1	0,5	91,39	0,75	0,01	92,15
2048	1	0,33	74,28	1,89	0,03	76,20
1536	1	0,2	55,80	3,69	0,02	59,50
1024	1	0,17	38,22	4,40	0,02	42,64

9.4.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,02	1	2.348,12	1,83	0,04	2.349,98	1	1	2.058,98	3,63	0,02	2.062,62
8192	0,02	1	1.796,53	1,58	0,02	1.798,12	1	0,73	1.585,85	4,31	0,04	1.590,19
6144	0,02	1	1.277,54	1,64	0,04	1.279,22	1	0,53	1.131,24	4,50	0,02	1.135,76
4096	0,02	1	721,86	1,58	0,02	723,46	1	0,1	626,97	4,00	0,04	631,01
3072	0,02	1	522,49	1,55	0,02	524,06	1	0,13	438,50	4,75	0,04	443,29
2560	0,02	1	408,52	1,66	0,02	410,20	1	0,13	341,45	4,89	0,04	346,38
2048	0,02	1	316,33	1,63	0,02	317,97	1	0,07	263,09	5,42	0,02	268,54
1536	0,02	1	239,25	2,55	0,04	241,84	1	0,07	198,81	5,65	0,02	204,48
1024	0,02	1	155,34	2,37	0,04	157,74	1	0,07	127,78	5,47	0,02	133,27

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,02	1	1.113,11	2,33	0,02	1.115,46	1	0,07	2.054,45	3,66	0,04	2.058,14
8192	0,02	1	769,50	1,94	0,02	771,45	1	0,07	1.575,01	3,61	0,02	1.578,64
6144	0,02	1	482,79	3,94	0,02	486,75	1	0,07	1.137,05	5,78	0,04	1.142,87
4096	0,02	1	257,53	4,89	0,02	262,44	1	0,07	627,79	5,16	0,02	632,97
3072	0,31	0,8	181,50	4,81	0,02	186,34	1	0,07	437,72	6,31	0,04	444,07
2560	0,28	0,83	146,74	5,58	0,04	152,36	1	0,07	337,38	6,63	0,02	344,03
2048	0,35	1	115,36	5,95	0,04	121,35	1	0,07	261,46	7,39	0,02	268,87
1536	0,39	1	87,00	6,84	0,02	93,86	1	0,07	194,77	7,19	0,02	201,98
1024	0,49	1	56,83	6,71	0,02	63,57	1	0,07	126,68	7,14	0,02	133,85

9.5 APÊNDICE 5: PRIMEIRO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO AES, COM CHAVES 192 BITS

9.5.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,17	260,69	4,84	0,02	265,543	1	0,17	256,91	4,83	0,02	261,76
8192	1	0,13	212,60	4,59	0,02	217,213	1	0,13	208,18	4,61	0,02	212,81
6144	1	0,10	156,81	5,67	0,04	162,518	1	0,10	154,00	5,61	0,02	159,62
4096	1	0,07	108,27	6,05	0,06	114,383	1	0,07	103,21	5,94	0,08	109,24
3072	1	0,07	81,63	5,99	0,06	87,68	1	0,07	77,60	6,03	0,06	83,69
2560	1	0,07	66,50	6,06	0,09	72,645	1	0,07	63,96	5,97	0,12	70,04
2048	1	0,07	56,02	6,26	0,09	62,369	1	0,07	53,42	6,09	0,06	59,58
1536	1	0,03	43,45	6,46	0,09	49,996	1	0,03	39,99	6,49	0,07	46,55
1024	1	0,03	30,44	6,42	0,27	37,14	1	0,03	26,45	6,10	0,06	32,61

	Jaccard						Overlap					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,17	253,10	4,87	0,02	257,98	1	0,17	259,67	4,78	0,02	264,47
8192	1	0,13	207,68	4,66	0,02	212,35	1	0,13	212,97	4,55	0,02	217,53
6144	1	0,10	153,96	5,55	0,04	159,54	1	0,10	157,52	5,59	0,02	163,13
4096	1	0,07	103,92	6,08	0,06	110,06	1	0,07	105,88	5,93	0,05	111,86
3072	1	0,07	77,50	5,98	0,09	83,57	1	0,07	78,74	6,10	0,07	84,91
2560	1	0,07	64,07	6,05	0,07	70,18	1	0,07	65,14	6,03	0,07	71,24
2048	1	0,07	53,08	6,13	0,08	59,28	1	0,07	54,74	6,19	0,07	61,01
1536	1	0,03	39,91	6,33	0,06	46,31	1	0,03	40,96	6,36	0,08	47,39
1024	1	0,03	28,02	6,12	0,05	34,19	1	0,03	27,64	6,15	0,07	33,87

	Simple-matching					
Tamanho dos textos (bytes)			Tempo			
	P	R	Index	Clus	Analis	total
10240	1	0,17	263,47	4,75	0,02	268,23
8192	1	0,13	212,38	4,56	0,02	216,96
6144	1	0,10	158,95	5,53	0,02	164,50
4096	1	0,07	106,25	5,93	0,02	112,21
3072	1	0,07	79,65	5,95	0,22	85,82
2560	1	0,07	66,80	6,00	0,07	72,87
2048	1	0,07	53,73	6,09	0,07	59,89
1536	1	0,03	40,55	6,28	0,07	46,89
1024	1	0,03	28,24	6,60	0,08	34,92

9.5.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,05	1	1.456,99	6,51	0,02	1.463,52	1	0,07	1.247,86	6,27	0,04	1.254,17
8192	0,05	1	1.115,24	6,23	0,04	1.121,50	1	0,07	945,78	6,21	0,02	952,01
6144	0,02	1	748,79	6,27	0,04	755,09	1	0,07	628,18	6,37	0,02	634,57
4096	0,05	1	455,34	6,46	0,02	461,82	1	0,07	376,07	6,17	0,04	382,27
3072	0,05	1	323,23	6,22	0,11	329,56	1	0,07	262,34	6,34	0,07	268,75
2560	0,05	1	261,16	6,65	0,08	267,90	1	0,07	214,07	6,41	0,11	220,59
2048	0,05	1	213,52	6,28	0,10	219,90	1	0,07	174,05	6,34	0,07	180,45
1536	0,05	1	156,09	6,44	0,07	162,60	1	0,03	118,86	6,51	0,06	125,44
1024	0,05	1	102,95	6,13	0,09	109,16	1	0,03	82,28	6,31	0,04	88,63

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,9	0,10	550,61	7,44	0,02	558,08	1	0,03	1.248,02	6,37	0,04	1.254,42
8192	0,94	0,10	390,26	8,19	0,04	398,49	1	0,03	951,43	6,34	0,04	957,81
6144	0,94	0,07	256,26	8,19	0,02	264,47	1	0,03	629,26	6,22	0,04	635,51
4096	0,94	0,07	159,76	7,98	0,04	167,78	1	0,03	373,87	6,34	0,02	380,24
3072	0,97	0,07	117,44	8,38	0,11	125,92	1	0,07	262,63	6,40	0,11	269,14
2560	0,97	0,07	95,17	8,59	0,10	103,86	1	0,03	214,44	6,74	0,06	221,24
2048	0,97	0,07	78,66	8,49	0,08	87,23	1	0,07	171,99	6,43	0,09	178,50
1536	0,97	0,07	56,99	8,53	0,06	65,58	1	0,03	124,94	6,42	0,09	131,45
1024	0,97	0,07	38,70	8,37	0,06	47,13	1	0,03	83,34	6,14	0,07	89,55

9.5.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)	P	R	Tempo				P	R	Tempo			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,07	260,69	5,34	0,02	266,05	1	0,07	256,91	5,32	0,02	262,25
8192	1	0,07	212,60	4,80	0,04	217,44	1	0,07	208,18	4,83	0,04	213,05
6144	1	0,07	156,81	5,66	0,02	162,49	1	0,07	154,00	5,62	0,04	159,66
4096	1	0,07	108,27	5,78	0,09	114,14	1	0,07	103,21	5,82	0,07	109,11

9.5.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Euclidiana	Manhattan

9.5.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,17	260,69	4,51	0,04	265,23	1	0,17	256,91	4,52	0,03	261,47
8192	1	0,13	212,60	4,31	0,02	216,92	1	0,13	208,18	4,28	0,02	212,48
6144	1	0,10	156,81	5,30	0,04	162,15	1	0,10	154,00	5,27	0,02	159,28
4096	1	0,07	108,27	5,66	0,09	114,01	1	0,07	103,21	5,66	0,11	108,98
3072	1	0,07	81,63	5,68	0,09	87,40	1	0,07	77,60	5,68	0,07	83,35
2560	1	0,07	66,50	5,66	0,06	72,22	1	0,07	63,96	5,69	0,10	69,75
2048	1	0,07	56,02	5,68	0,08	61,78	1	0,07	53,42	5,68	0,07	59,17
1536	1	0,03	43,45	5,83	0,09	49,36	1	0,03	39,99	5,90	0,08	45,98
1024	1	0,03	30,44	5,84	0,07	36,35	1	0,03	26,45	5,81	0,06	32,32

	Jaccard						Overlap					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,17	253,10	4,51	0,03	257,64	1	0,17	259,67	4,48	0,03	264,19
8192	1	0,13	207,68	4,30	0,02	211,99	1	0,13	212,97	4,24	0,02	217,22
6144	1	0,10	153,96	5,29	0,04	159,28	1	0,10	157,52	5,28	0,02	162,82
4096	1	0,07	103,92	5,66	0,02	109,60	1	0,07	105,88	5,65	0,04	111,57
3072	1	0,07	77,50	5,71	0,09	83,30	1	0,07	78,74	5,67	0,07	84,48
2560	0,07	0,04	64,07	5,69	0,09	69,84	1	0,07	65,14	5,65	0,06	70,84
2048	1	0,07	53,08	5,69	0,08	58,85	1	0,07	54,74	5,67	0,06	60,47
1536	1	0,03	39,91	5,84	0,06	45,82	1	0,03	40,96	5,81	0,06	46,82
1024	1	0,03	28,02	5,89	0,07	33,99	1	0,03	27,64	5,81	0,06	33,51

	Simple-matching					
Tamanho dos textos (bytes)			Tempo			
	P	R	Index	Clus	Analis	total
10240	1	0,17	263,47	4,48	0,02	267,97
8192	1	0,13	212,38	4,29	0,03	216,70
6144	1	0,10	158,95	5,31	0,02	164,27
4096	1	0,07	106,25	5,63	0,02	111,90
3072	1	0,07	79,65	5,70	0,10	85,46
2560	1	0,07	66,80	5,70	0,07	72,58
2048	1	0,67	53,73	5,62	0,08	59,43
1536	1	0,03	40,55	5,84	0,06	46,45
1024	1	0,03	28,24	5,86	0,07	34,17

9.5.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,02	1	1.456,99	1,46	0,02	1.458,47	1	0,07	1.247,86	5,97	0,05	1.253,87
8192	0,02	1	1.115,24	0,99	0,04	1.116,27	1	0,07	945,78	5,79	0,02	951,58
6144	0,02	1	748,79	1,13	0,04	749,95	1	0,07	628,18	5,84	0,02	634,05
4096	0,02	1	455,34	1,03	0,02	456,39	1	0,07	376,07	6,03	0,02	382,12
3072	0,02	1	323,23	0,87	0,10	324,20	1	0,07	262,34	5,99	0,10	268,44
2560	0,02	1	261,16	0,85	0,06	262,08	1	0,07	214,07	6,05	0,07	220,20
2048	0,02	1	213,52	0,70	0,06	214,28	1	0,07	174,05	6,10	0,07	180,22
1536	0,02	1	156,09	6,01	0,07	162,17	1	0,03	118,86	6,03	0,07	124,97
1024	0,02	1	102,95	5,79	0,06	108,80	1	0,03	82,28	5,99	0,08	88,35

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,82	1	550,61	6,38	0,02	557,01	1	0,03	1.248,02	6,92	0,04	1.254,98
8192	0,82	1	390,26	7,60	0,02	397,89	1	0,03	951,43	6,64	0,04	958,11
6144	0,85	1	256,26	7,80	0,02	264,07	1	0,03	629,26	7,66	0,04	636,96
4096	0,85	1	159,76	7,72	0,04	167,52	1	0,03	373,87	6,16	0,04	380,07
3072	0,89	1	117,44	8,13	0,11	125,67	1	0,07	262,63	8,14	0,14	270,92
2560	0,89	1	95,17	8,24	0,08	103,49	1	0,03	214,44	6,40	0,09	220,93
2048	0,89	1	78,66	8,10	0,08	86,84	1	0,07	171,99	6,43	0,09	178,51
1536	0,89	1	56,99	8,09	0,08	65,16	1	0,03	124,94	7,97	0,09	133,00
1024	0,89	1	38,70	8,17	0,07	46,93	1	0,03	83,34	5,85	0,06	89,24

9.6 APÊNDICE 6: PRIMEIRO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO AES, COM CHAVES 256 BITS

9.6.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)	P	R	Tempo				P	R	Tempo			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,23	186,41	3,18	0,02	189,62	1	0,23	183,33	3,06	0,02	186,42
8192	1	0,17	147,22	3,99	0,03	151,24	1	0,17	143,84	3,88	0,03	147,75
6144	1	0,10	110,37	5,16	0,04	115,57	1	0,10	106,45	5,15	0,05	111,64
4096	1	0,03	74,70	6,24	0,10	81,04	1	0,03	72,68	6,21	0,08	78,97
3072	1	0,03	57,75	6,30	0,09	64,14	1	0,03	54,42	6,30	0,07	60,79
2560	1	0,03	49,96	6,80	0,10	56,85	1	0,03	46,75	6,66	0,08	53,48
2048	1	0,03	39,72	6,50	0,09	46,31	1	0,03	38,12	6,43	0,09	44,64
1536	1	0,03	29,11	6,53	0,11	35,74	1	0,03	28,77	6,63	0,07	35,46
1024	1	0,03	20,76	7,58	0,19	28,53	1	0,03	20,35	6,87	0,07	27,29

	Jaccard						Overlap					
Tamanho dos textos (bytes)	P	R	Tempo				P	R	Tempo			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,23	180,48	3,40	0,04	183,92	1	0,23	185,27	3,03	0,04	188,34
8192	1	0,17	142,71	3,88	0,02	146,62	1	0,17	146,17	3,87	0,04	150,08
6144	1	0,10	106,22	5,08	0,03	111,33	1	0,10	108,01	5,31	0,03	113,34
4096	1	0,03	73,24	6,31	0,10	79,65	1	0,03	73,49	6,34	0,05	79,88
3072	1	0,03	54,54	6,24	0,08	60,86	1	0,03	55,15	6,29	0,09	61,53
2560	1	0,03	46,97	6,72	0,07	53,76	1	0,03	47,20	6,44	0,07	53,71
2048	1	0,03	37,87	6,67	0,08	44,61	1	0,03	38,31	6,35	0,08	44,74
1536	1	0,03	28,65	6,49	0,09	35,23	1	0,03	29,60	6,61	0,08	36,28
1024	1	0,03	20,81	6,63	0,09	27,53	1	0,03	20,68	6,67	0,11	27,46

	Simple-matching					
Tamanho dos textos (bytes)	P	R	Tempo			
			Index	Clus	Analis	total
10240	1	0,23	186,74	3,17	0,02	189,93
8192	1	0,17	147,71	3,81	0,03	151,54
6144	1	0,10	107,32	5,30	0,03	112,65
4096	1	0,03	74,94	6,23	0,05	81,21
3072	1	0,03	54,73	6,27	0,10	61,09
2560	1	0,03	47,18	6,39	0,10	53,67
2048	1	0,03	37,92	6,50	0,08	44,50
1536	1	0,03	29,50	6,49	0,07	36,06
1024	1	0,03	20,11	6,60	0,10	26,81

9.6.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,05	1	1015,31	8,43	0,04	1.023,79	1	0,07	847,70	6,20	0,03	853,93
8192	0,05	1	757,37	6,59	0,04	764,00	1	0,07	629,39	6,43	0,03	635,84
6144	0,05	1	510,74	6,42	0,03	517,19	1	0,07	432,01	6,33	0,05	438,38
4096	0,05	1	328,90	6,54	0,04	335,49	1	0,03	264,04	6,40	0,03	270,47
3072	0,05	1	234,85	6,38	0,11	241,33	1	0,03	195,53	6,49	0,09	202,11
2560	0,05	1	196,88	6,84	0,05	203,77	1	0,03	159,86	6,54	0,07	166,47
2048	0,05	1	156,30	6,08	0,07	162,45	1	0,03	127,44	6,45	0,09	133,97
1536	0,05	1	108,36	6,75	0,07	115,19	1	0,03	91,88	6,58	0,08	98,54
1024	0,05	1	76,52	6,81	0,09	83,43	1	0,03	59,70	6,70	0,08	66,47

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,87	0,13	344,30	7,41	0,05	351,75	1	0,03	860,62	6,31	0,07	867,00
8192	0,94	0,07	253,50	7,37	0,05	260,92	1	0,03	626,76	6,21	0,04	633,02
6144	0,94	0,07	179,75	8,31	0,05	188,11	1	0,07	434,30	6,64	0,03	440,97
4096	0,94	0,07	116,84	8,20	0,03	125,07	1	0,03	264,29	6,37	0,03	270,69
3072	0,97	0,07	86,43	8,55	0,08	95,06	1	0,03	195,39	6,37	0,10	201,85
2560	0,97	0,07	72,23	8,85	0,06	81,14	1	0,03	160,98	6,54	0,08	167,60
2048	0,97	0,07	56,62	8,60	0,10	65,32	1	0,03	127,78	6,53	0,07	134,39
1536	0,97	0,07	42,12	8,72	0,07	50,91	1	0,03	91,12	6,56	0,08	97,76
1024	0,97	0,07	28,36	8,93	0,05	37,33	1	0,03	60,42	6,61	0,05	67,08

9.6.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,07	186,41	4,49	0,03	190,92	1	0,07	183,33	4,47	0,03	187,83
8192	1	0,07	147,22	5,00	0,04	152,26	1	0,07	143,84	5,00	0,03	148,87
6144	1	0,07	110,37	5,66	0,03	116,05	1	0,07	106,45	5,66	0,03	112,14
4096	1	0,03	74,70	5,94	0,09	80,73	1	0,03	72,68	6,02	0,09	78,79
3072	1	0,03	57,75	5,96	0,08	63,79	1	0,03	54,42	5,97	0,10	60,49
2560	1	0,03	49,96	5,95	0,08	55,99	1	0,03	46,75	5,99	0,09	52,82
2048	1	0,03	39,72	5,96	0,08	45,76	1	0,03	38,12	5,98	0,06	44,16
1536	1	0,03	29,11	5,96	0,06	35,13	1	0,03	28,77	5,96	0,07	34,80
1024	1	0,03	20,76	6,05	0,07	26,87	1	0,03	20,35	5,99	0,09	26,43

Tamanho dos textos (bytes)	Jaccard						Overlap					
			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,07	180,48	4,50	0,03	185,00	1	0,07	185,27	4,50	0,03	189,80
8192	1	0,07	142,71	4,99	0,03	147,72	1	0,07	146,17	4,98	0,03	151,18
6144	1	0,07	106,22	5,67	0,05	111,94	1	0,07	108,01	5,65	0,05	113,71
4096	1	0,03	73,24	6,00	0,10	79,34	1	0,03	73,49	5,94	0,03	79,46
3072	1	0,03	54,54	5,98	0,09	60,61	1	0,03	55,15	5,98	0,08	61,22
2560	1	0,03	46,97	5,97	0,08	53,02	1	0,03	47,20	5,98	0,09	53,27
2048	1	0,03	37,87	5,98	0,07	43,92	1	0,03	38,31	5,97	0,09	44,37
1536	1	0,03	28,65	5,94	0,09	34,68	1	0,03	29,60	5,94	0,09	35,62
1024	1	0,03	20,81	5,99	0,08	26,88	1	0,03	20,68	6,00	0,06	26,74

Tamanho dos textos (bytes)	Simple-matching					
			Tempo			
	P	R	Index	Clus	Analis	total
10240	1	0,07	186,74	4,50	0,04	191,28
8192	1	0,07	147,71	4,97	0,04	152,72
6144	1	0,07	107,32	5,68	0,05	113,05
4096	1	0,03	74,94	5,93	0,03	80,89
3072	1	0,03	54,73	5,93	0,07	60,73
2560	1	0,03	47,18	5,99	0,07	53,24
2048	1	0,03	37,92	5,97	0,11	44,00
1536	1	0,03	29,50	5,99	0,10	35,58
1024	1	0,03	20,11	5,98	0,07	26,17

9.6.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,02	1	1015,31	5,82	0,03	1.021,16	1	0,23	847,70	5,64	0,05	853,39
8192	0,02	1	757,37	5,68	0,03	763,08	1	0,17	629,39	5,41	0,05	634,84
6144	0,02	1	510,74	6,18	0,05	516,97	1	0,10	432,01	5,82	0,03	437,86
4096	0,02	1	328,90	6,15	0,03	335,07	1	0,03	264,04	6,02	0,03	270,09
3072	0,02	1	234,85	6,04	0,08	240,97	1	0,03	195,53	6,06	0,09	201,67
2560	0,02	1	196,88	6,06	0,09	203,03	1	0,03	159,86	6,04	0,08	165,97
2048	0,02	1	156,30	6,02	0,08	162,41	1	0,03	127,44	6,01	0,07	133,52
1536	0,02	1	108,36	6,03	0,07	114,46	1	0,03	91,88	6,02	0,08	97,97
1024	0,02	1	76,52	6,08	0,08	82,68	1	0,03	59,70	6,01	0,09	65,79

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,02	1	344,30	6,41	0,03	350,74	1	0,03	860,62	5,66	0,05	866,33
8192	0,02	1	253,50	6,43	0,03	259,96	1	0,03	626,76	5,41	0,03	632,20
6144	0,02	1	179,75	7,68	0,03	187,46	1	0,07	434,30	5,85	0,03	440,19
4096	0,02	1	116,84	7,74	0,05	124,62	1	0,03	264,29	6,08	0,05	270,41
3072	0,02	1	86,43	8,24	0,11	94,78	1	0,03	195,39	6,02	0,09	201,50
2560	0,02	1	72,23	8,20	0,11	80,54	1	0,03	160,98	6,04	0,08	167,11
2048	0,02	1	56,62	8,11	0,08	64,81	1	0,03	127,78	6,01	0,10	133,89
1536	0,02	1	42,12	8,10	0,07	50,29	1	0,03	91,12	6,03	0,09	97,24
1024	0,02	1	28,36	8,22	0,10	36,69	1	0,03	60,42	6,01	0,10	66,52

9.6.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)	P	R	Tempo				P	R	Tempo			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,23	186,41	2,79	0,04	189,24	1	0,23	183,33	2,79	0,02	186,14

9.6.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,02	1	1015,31	1,16	0,04	1.016,52	1	0,07	847,70	5,60	0,03	853,33
8192	0,02	1	757,37	1,50	0,04	758,91	1	0,07	629,39	5,40	0,03	634,81
6144	0,02	1	510,74	1,48	0,03	512,25	1	0,07	432,01	5,81	0,03	437,85
4096	0,02	1	328,90	5,94	0,05	334,88	1	0,03	264,04	6,01	0,04	270,09
3072	0,02	1	234,85	5,75	0,11	240,70	1	0,03	195,53	6,01	0,08	201,62
2560	0,02	1	196,88	6,05	0,07	203,00	1	0,03	159,86	6,00	0,09	165,95
2048	0,02	1	156,30	5,88	0,07	162,25	1	0,03	127,44	6,04	0,07	133,54
1536	0,02	1	108,36	6,04	0,10	114,50	1	0,03	91,88	6,01	0,10	97,99
1024	0,02	1	76,52	5,98	0,09	82,60	1	0,03	59,70	6,00	0,08	65,78

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo						Tempo			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,77	0,27	344,30	6,27	0,05	350,62	1	0,03	860,62	6,70	0,03	867,36
8192	0,84	0,23	253,50	6,37	0,03	259,89	1	0,03	626,76	6,77	0,05	633,57
6144	0,84	0,20	179,75	7,93	0,05	187,72	1	0,07	434,30	7,27	0,04	441,61
4096	0,89	1	116,84	7,70	0,03	124,57	1	0,03	264,29	5,98	0,03	270,29
3072	0,92	1	86,43	8,13	0,08	94,64	1	0,03	195,39	7,92	0,08	203,38
2560	0,92	1	72,23	8,14	0,08	80,45	1	0,03	160,98	7,54	0,08	168,60
2048	0,92	1	56,62	8,09	0,11	64,82	1	0,03	127,78	7,92	0,09	135,79
1536	0,92	1	42,12	8,08	0,10	50,30	1	0,03	91,12	5,69	0,07	96,88
1024	0,92	1	28,36	8,15	0,07	36,58	1	0,03	60,42	7,10	0,09	67,60

9.7 APÊNDICE 7: PRIMEIRO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO RSA, COM CHAVES 64 BITS

9.7.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	1	750,11	0,93	0,01	751,05	1	1	778,13	0,96	0,01	779,09
8192	1	1	618,62	0,97	0,01	619,59	1	1	610,32	1,02	0,01	611,34
6144	1	1	458,36	1,00	0,01	459,36	1	1	448,46	1,08	0,02	449,53
4096	1	1	302,70	1,02	0,02	303,72	1	1	299,01	1,06	0,01	300,07
2048	1	1	167,32	0,79	0,07	168,11	1	1	160,03	0,81	0,08	160,85
1024	1	1	84,25	0,87	0,05	85,13	1	1	82,04	0,94	0,06	82,98
512	1	1	45,36	0,86	0,06	46,22	1	1	43,50	0,84	0,06	44,34
256	1	1	24,14	0,81	0,09	24,95	1	1	23,46	1,81	0,05	25,28
192	1	0,5	19,40	0,84	0,07	20,24	1	0,5	18,36	0,81	0,06	19,16
128	1	0,1	14,36	1,64	0,05	16,00	1	0,1	13,65	1,66	0,06	15,31
64	1	0,2	10,28	1,19	0,16	11,47	1	0,2	8,94	0,97	0,06	9,91

Tamanho dos textos (bytes)	Jaccard						Overlap					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	1	749,80	1,04	0,01	750,84	1	1	752,67	0,91	0,01	753,58
8192	1	1	615,78	0,98	0,01	616,77	1	1	614,99	0,83	0,01	615,83
6144	1	1	454,73	0,99	0,01	455,72	1	1	457,34	0,94	0,01	458,28
4096	1	1	299,60	1,01	0,01	300,61	1	1	302,45	1,03	0,01	303,48
2048	1	1	161,27	0,81	0,07	162,07	1	1	163,19	0,80	0,06	163,99
1024	1	1	80,47	1,01	0,08	81,48	1	1	82,64	0,87	0,06	83,50
512	1	1	42,96	0,99	0,07	43,96	1	1	44,24	0,98	0,07	45,22
256	1	1	23,37	0,80	0,08	24,16	1	1	23,89	0,79	0,06	24,67
192	1	0,50	18,49	0,88	0,07	19,37	1	0,50	18,98	0,83	0,07	19,81
128	1	0,10	14,66	4,68	0,05	19,34	1	0,10	14,34	1,63	0,06	15,97
64	1	0,20	8,86	0,91	0,08	9,77	1	0,20	9,29	0,92	0,07	10,21

Tamanho dos textos (bytes)	Simple-matching					
	P	R	Tempo (segundos)			
			Index	Clus	Analis	total
10240	1	1	760,34	0,94	0,01	761,28
8192	1	1	619,33	0,93	0,01	620,26
6144	1	1	459,64	0,93	0,01	460,58
4096	1	1	300,82	0,92	0,01	301,74
2048	1	1	160,20	0,97	0,01	161,17
1024	1	1	82,43	0,79	0,08	83,22
512	1	1	43,41	0,80	0,09	44,22
256	1	1	23,12	1,10	0,05	24,22
192	1	0,50	18,42	0,90	0,09	19,31
128	1	0,10	14,39	1,59	0,06	15,98
64	1	0,20	8,61	0,97	0,06	9,58

9.7.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	5.305,30	5,63	0,02	5.310,93	0,02	1	4.634,41	6,35	0,01	4.640,76
8192	1	0,03	4.251,95	6,17	0,04	4.258,12	1	0,03	3.681,88	6,49	0,02	3.688,37
6144	1	0,03	3.082,71	6,23	0,02	3.088,94	1	0,03	2.660,53	6,42	0,04	2.666,95
4096	1	0,03	1.998,04	6,05	0,02	2.004,09	1	0,03	1.726,85	6,39	0,04	1.733,24
2048	1	0,03	868,55	6,75	0,09	875,31	1	0,03	728,21	6,40	0,08	734,61
1024	1	0,03	366,29	6,71	0,10	373,00	1	0,03	296,14	6,17	0,07	302,30
512	1	0,03	174,55	6,72	0,11	181,27	1	0,03	139,58	6,30	0,07	145,87
256	1	0,03	85,36	7,66	0,09	93,02	1	0,03	67,98	6,21	0,19	74,19
192	1	0,03	63,59	6,30	0,10	69,89	1	0,03	49,81	6,38	0,10	56,19
128	1	0,03	43,60	6,26	0,09	49,86	1	0,03	33,53	6,16	0,08	39,69
64	1	0,03	22,25	8,30	0,08	30,55	1	0,03	17,52	6,24	0,06	23,77

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,18	0,04	2.680,85	5,71	0,01	2.686,56	1	0,03	4.631,50	1,11	0,02	4.632,62
8192	1	0,03	2.146,07	6,05	0,04	2.152,12	1	0,03	3.723,07	1,12	0,02	3.724,19
6144	1	0,03	1.496,90	6,08	0,02	1.502,98	1	0,03	2.680,34	0,98	0,04	2.681,32
4096	1	0,03	835,12	6,44	0,02	841,56	1	0,03	1.744,73	1,16	0,04	1.745,88
2048	1	0,03	284,45	6,57	0,09	291,02	1	0,03	724,86	1,08	0,02	725,94
1024	1	0,03	130,73	6,51	0,07	137,24	1	0,03	295,54	0,94	0,09	296,48
512	1	0,03	65,06	7,10	0,09	72,17	1	0,03	139,15	1,16	0,05	140,30
256	1	0,03	32,87	7,36	0,09	40,22	1	0,03	67,77	5,73	0,05	73,50
192	1	0,03	24,96	8,11	0,09	33,06	1	0,03	49,33	6,17	0,08	55,50
128	1	0,03	17,41	8,35	0,08	25,76	1	0,03	33,73	5,95	0,07	39,68
64	1	0,03	10,08	8,94	0,09	19,02	1	0,03	17,21	6,40	0,07	23,61

9.7.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	1	750,11	0,71	0,01	750,84	1	1	778,13	0,71	0,02	778,85
8192	1	1	618,62	0,71	0,01	619,35	1	1	610,32	0,73	0,03	611,08
6144	1	1	458,36	0,71	0,03	459,10	1	1	448,46	0,70	0,01	449,17
4096	1	1	302,70	0,71	0,01	303,42	1	1	299,01	0,71	0,02	299,73
2048	1	1	167,32	0,71	0,06	168,08	1	1	160,03	0,72	0,06	160,82
1024	1	0,25	84,25	0,81	0,07	85,13	1	0,25	82,04	0,81	0,08	82,93
512	1	0,13	45,36	1,15	0,05	46,56	1	0,13	43,50	1,13	0,08	44,72
256	1	0,07	24,14	2,26	0,08	26,48	1	0,07	23,46	2,26	0,07	25,80
192	1	0,59	19,40	2,83	0,07	22,30	1	0,59	18,36	2,86	0,08	21,30
128	1	0,06	14,36	2,53	0,08	16,96	1	0,06	13,65	2,57	0,06	16,28
64	1	0,07	10,28	2,36	0,07	12,71	1	0,07	8,94	2,43	0,09	11,46

Tamanho dos textos (bytes)	Jaccard						Overlap					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	1	749,80	0,71	0,03	750,54	1	1	752,67	0,71	0,01	753,40
8192	1	1	615,78	0,72	0,02	616,52	1	1	614,99	0,71	0,01	615,71
6144	1	1	454,73	0,71	0,01	455,46	1	1	457,34	0,71	0,03	458,08
4096	1	1	299,60	0,72	0,01	300,34	1	1	302,45	0,71	0,03	303,19
2048	1	1	161,27	0,71	0,07	162,04	1	1	163,19	0,71	0,06	163,96
1024	1	0,25	80,47	0,81	0,08	81,36	1	0,33	82,64	0,75	0,07	83,46
512	1	0,13	42,96	1,13	0,09	44,18	1	0,11	44,24	1,30	0,07	45,61
256	1	0,07	23,37	2,26	0,06	25,69	1	0,07	23,89	2,37	0,06	26,32
192	1	0,59	18,49	2,86	0,06	21,41	1	0,06	18,98	2,66	0,08	21,73
128	1	0,06	14,66	2,55	0,08	17,29	1	0,06	14,34	2,55	0,06	16,95
64	1	0,07	8,86	2,35	0,07	11,27	1	0,07	9,29	2,33	0,07	11,69

Tamanho dos textos (bytes)	Simple-matching					
	P	R	Tempo (segundos)			
			Index	Clus	Analis	total
10240	1	1	760,34	0,71	0,03	761,08
8192	1	1	619,33	0,72	0,02	620,06
6144	1	1	459,64	0,69	0,02	460,35
4096	1	1	300,82	0,71	0,01	301,55
2048	1	1	160,20	0,71	0,02	160,92
1024	1	0,33	82,43	0,75	0,09	83,27
512	1	0,14	43,41	0,99	0,10	44,50
256	1	0,08	23,12	1,98	0,07	25,18
192	1	0,06	18,42	2,67	0,07	21,15
128	1	0,06	14,39	2,57	0,05	17,01
64	1	0,07	8,61	2,35	0,06	11,02

9.7.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	5.305,30	5,33	0,04	5.310,67	0,02	0,17	4.634,41	5,95	0,02	4.640,39
8192	1	0,03	4.251,95	6,08	0,02	4.258,05	1	0,03	3.681,88	6,04	0,04	3.687,96
6144	1	0,03	3.082,71	5,81	0,04	3.088,56	1	0,03	2.660,53	9,07	0,02	2.669,62
4096	1	0,03	1.998,04	5,61	0,04	2.003,69	1	0,03	1.726,85	6,02	0,02	1.732,89
2048	1	0,03	868,55	6,46	0,10	875,11	1	0,03	728,21	5,97	0,10	734,28
1024	1	0,03	366,29	6,25	0,11	372,64	1	0,03	296,14	5,79	0,09	302,01
512	1	0,03	174,55	6,19	0,06	180,81	1	0,03	139,58	5,09	0,09	144,75
256	1	0,03	85,36	5,95	0,07	91,38	1	0,03	67,98	5,01	0,09	73,08
192	1	0,03	63,59	5,15	0,07	68,81	1	0,03	49,81	5,20	0,09	55,10
128	1	0,03	43,60	4,82	0,07	48,49	1	0,03	33,53	5,19	0,07	38,79
64	1	0,03	22,25	4,69	0,06	26,99	1	0,03	17,52	4,61	0,09	22,22

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	0,58	0,03	2.680,85	5,59	0,03	2.686,48	1	0,03	4.631,50	0,75	0,04	4.632,28
8192	1	0,03	2.146,07	5,77	0,02	2.151,87	1	0,03	3.723,07	0,74	0,04	3.723,84
6144	1	0,03	1.496,90	5,96	0,04	1.502,90	1	0,03	2.680,34	0,80	0,02	2.681,16
4096	1	0,03	835,12	6,05	0,04	841,21	1	0,03	1.744,73	0,80	0,02	1.745,55
2048	1	0,03	284,45	6,42	0,02	290,89	1	0,03	724,86	1,26	0,02	726,14
1024	1	0,03	130,73	6,07	0,10	136,90	1	0,03	295,54	1,70	0,09	297,32
512	1	0,03	65,06	5,44	0,06	70,57	1	0,03	139,15	3,31	0,11	142,56
256	1	0,03	32,87	5,63	0,06	38,55	1	0,03	67,77	4,00	0,09	71,86
192	1	0,03	24,96	6,07	0,08	31,11	1	0,03	49,33	5,18	0,07	54,58
128	1	0,03	17,41	6,38	0,07	23,87	1	0,03	33,73	5,19	0,08	39,00
64	1	0,03	10,08	4,62	0,06	14,76	1	0,03	17,21	4,61	0,07	21,90

9.7.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	1	750,11	0,62	0,01	750,75	1	1	778,13	0,64	0,01	778,78
8192	1	1	618,62	0,62	0,01	619,25	1	1	610,32	0,63	0,03	610,98
6144	1	1	458,36	0,61	0,01	458,99	1	1	448,46	0,62	0,01	449,09
4096	1	1	302,70	0,62	0,03	303,35	1	1	299,01	0,61	0,01	299,64
2048	1	1	167,32	0,62	0,08	168,01	1	1	160,03	0,61	0,06	160,71
1024	1	1	84,25	0,62	0,06	84,93	1	1	82,04	0,63	0,08	82,74
512	1	1	45,36	0,62	0,08	46,07	1	1	43,50	0,61	0,08	44,19
256	1	1	24,14	0,59	0,05	24,78	1	1	23,46	0,58	0,06	24,10
192	1	0,5	19,40	0,59	0,08	20,07	1	0,50	18,36	0,59	0,08	19,02
128	1	0,1	14,36	1,36	0,06	15,77	1	0,10	13,65	1,37	0,08	15,09

9.7.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,03	5.305,30	2,63	0,10	5.308,03	0,02	1	4.634,41	5,96	0,01	4.640,38
8192	1	0,03	4.251,95	4,18	0,04	4.256,17	1	0,03	3.681,88	5,97	0,02	3.687,87
6144	1	0,03	3.082,71	3,05	0,03	3.085,79	1	0,03	2.660,53	5,99	0,04	2.666,56
4096	1	0,03	1.998,04	3,56	0,02	2.001,62	1	0,03	1.726,85	5,96	0,04	1.732,85
2048	1	0,03	868,55	2,89	0,08	871,53	1	0,03	728,21	5,96	0,08	734,25
1024	1	0,03	366,29	1,83	0,09	368,21	1	0,03	296,14	5,76	0,09	301,99
512	1	0,03	174,55	1,42	0,08	176,06	1	0,03	139,58	7,26	0,07	146,91
256	1	0,03	85,36	1,71	0,09	87,16	1	0,03	67,98	4,37	0,06	72,41
192	1	0,03	63,59	1,49	0,10	65,18	1	0,03	49,81	5,12	0,08	55,01
128	1	0,03	43,60	0,99	0,07	44,66	1	0,03	33,53	5,18	0,07	38,77
64	1	0,03	22,25	6,71	0,09	29,04	1	0,03	17,52	4,58	0,07	22,17

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	0,51	0,03	2.680,85	3,02	0,02	2.683,90	1	0,03	4.631,50	0,75	0,02	4.632,27
8192	1	0,03	2.146,07	3,30	0,04	2.149,41	1	0,03	3.723,07	0,75	0,02	3.723,85
6144	1	0,03	1.496,90	2,95	0,02	1.499,87	1	0,03	2.680,34	0,76	0,04	2.681,13
4096	1	0,03	835,12	2,68	0,02	837,82	1	0,03	1.744,73	0,73	0,04	1.745,50
2048	1	0,03	284,45	1,95	0,04	286,44	1	0,03	724,86	0,76	0,04	725,65
1024	1	0,03	130,73	1,78	0,06	132,57	1	0,03	295,54	0,81	0,09	296,43
512	1	0,03	65,06	3,19	0,10	68,35	1	0,03	139,15	1,41	0,09	140,65
256	1	0,03	32,87	5,08	0,08	38,02	1	0,03	67,77	5,53	0,09	73,39
192	1	0,03	24,96	6,12	0,10	31,17	1	0,03	49,33	6,54	0,07	55,95
128	1	0,03	17,41	6,39	0,08	23,88	1	0,03	33,73	6,23	0,08	40,05
64	1	0,03	10,08	4,57	0,09	14,74	1	0,03	17,21	7,40	0,07	24,69

9.8 APÊNDICE 8 PRIMEIRO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO RSA, COM CHAVES 128 BITS

9.8.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	1	397,23	1,00	0,01	398,24	1	1	394,51	1,03	0,01	395,55
8192	1	1	316,49	1,00	0,01	317,51	1	1	314,88	0,98	0,01	315,88
6144	1	1	231,28	1,06	0,01	232,35	1	1	230,97	1,01	0,01	231,99
4096	1	1	158,87	1,01	0,04	159,92	1	1	156,02	0,94	0,04	157,00
2048	1	0,20	79,84	1,22	0,02	81,08	1	0,20	77,73	1,19	0,03	78,95
1024	1	0,04	42,94	5,05	0,08	48,07	1	0,04	41,18	4,98	0,07	46,22
512	1	0,04	22,99	8,90	0,07	31,95	1	0,04	22,14	6,29	0,09	28,51
256	1	0,07	13,03	3,17	0,03	16,23	1	0,07	12,80	3,29	0,04	16,13
192	1	0,03	10,23	7,23	0,03	17,48	1	0,03	10,01	7,53	0,06	17,61
128	1	0,03	8,25	7,79	0,07	16,12	1	0,03	8,09	7,91	0,05	16,05
64	1	0,03	6,67	9,42	0,32	16,41	1	0,03	5,89	8,11	0,10	14,10

Tamanho dos textos (bytes)	Jaccard						Overlap					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	1	394,25	0,96	0,01	395,22	1	1	396,03	1,09	0,01	397,14
8192	1	1	312,87	0,99	0,04	313,89	1	1	313,60	1,00	0,01	314,61
6144	1	1	231,10	0,97	0,01	232,08	1	1	231,91	1,07	0,03	233,01
4096	1	1	156,54	1,02	0,03	157,60	1	1	156,66	0,94	0,02	157,62
2048	1	0,20	77,94	1,15	0,03	79,11	1	0,20	79,20	1,13	0,05	80,37
1024	1	0,04	41,15	4,93	0,09	46,17	1	0,04	41,38	4,99	0,11	46,48
512	1	0,04	22,18	6,70	0,07	28,95	1	0,04	22,40	5,88	0,08	28,36
256	1	0,07	12,80	3,55	0,03	16,38	1	0,07	13,07	3,75	0,04	16,86
192	1	0,03	9,95	7,38	0,03	17,36	1	0,03	10,16	7,55	0,04	17,76
128	1	0,03	8,01	7,88	0,08	15,97	1	0,03	8,15	7,99	0,03	16,16
64	1	0,03	5,86	7,99	0,07	13,92	1	0,03	5,98	7,99	0,08	14,04

Tamanho dos textos (bytes)	Simple-matching					
	P	R	Tempo (segundos)			
			Index	Clus	Analis	total
10240	1	1	400,24	0,92	0,02	401,17
8192	1	1	320,79	0,91	0,01	321,71
6144	1	1	237,05	1,01	0,03	238,09
4096	1	1	158,43	0,96	0,02	159,41
2048	1	0,20	79,44	1,13	0,02	80,59
1024	1	0,04	40,87	4,96	0,04	45,86
512	1	0,04	22,02	5,81	0,10	27,92
256	1	0,07	12,62	2,84	0,09	15,55
192	1	0,03	10,04	7,14	0,03	17,21
128	1	0,03	7,76	7,37	0,03	15,16
64	1	0,03	6,66	7,83	0,06	14,56

9.8.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,03	2.484,13	6,43	0,02	2.490,58	1	0,03	2.139,14	6,18	0,04	2.145,36
8192	1	0,03	1.938,28	6,82	0,04	1.945,13	1	0,03	1.673,96	6,41	0,04	1.680,41
6144	1	0,03	1.384,48	6,99	0,04	1.391,50	1	0,03	1.180,16	6,54	0,02	1.186,72
4096	1	0,03	815,81	6,23	0,03	822,08	1	0,03	686,60	6,55	0,04	693,19
2048	1	0,03	350,44	7,44	0,04	357,92	1	0,03	287,03	6,47	0,03	293,53
1024	1	0,03	168,35	6,65	0,04	175,04	1	0,03	137,71	6,48	0,03	144,23
512	1	0,03	83,05	6,49	0,08	89,62	1	0,03	65,06	6,61	0,09	71,76
256	1	0,03	41,60	7,20	0,10	48,90	1	0,03	32,16	7,07	0,07	39,30
192	1	0,03	28,69	8,11	0,04	36,83	1	0,03	21,82	7,91	0,03	29,76
128	1	0,03	19,39	7,80	0,03	27,22	1	0,03	15,36	7,87	0,04	23,27
64	1	0,03	11,08	7,80	0,06	18,94	1	0,03	9,62	8,03	0,10	17,75

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,03	1.185,80	9,73	0,05	1.195,59	1	0,03	2.163,16	1,15	0,02	2.164,33
8192	1	0,03	840,01	7,23	0,04	847,27	1	0,03	1.685,02	3,62	0,02	1.688,66
[redacted]											0,04	1.198,07

9.8.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,20	397,23	0,90	0,02	398,14	1	0,20	394,51	0,87	0,01	395,39
8192	1	0,10	316,49	1,48	0,02	317,99	1	0,10	314,88	1,47	0,02	316,37
6144	1	0,08	231,28	2,03	0,02	233,33	1	0,08	230,97	2,07	0,02	233,06
4096	1	0,08	158,87	1,87	0,02	160,76	1	0,08	156,02	1,87	0,02	157,91
2048	1	0,06	79,84	2,77	0,03	82,64	1	0,06	77,73	2,79	0,05	80,58
1024	1	0,04	42,94	5,34	0,07	48,35	1	0,04	41,18	5,26	0,09	46,53
512	1	0,04	22,99	5,42	0,13	28,53	1	0,04	22,14	5,43	0,09	27,65
256	1	0,07	13,03	2,13	0,04	15,21	1	0,07	12,80	2,15	0,04	14,99
192	1	0,03	10,23	5,81	0,04	16,08	1	0,03	10,01	5,86	0,04	15,91
128	1	0,03	8,25	6,00	0,09	14,34	1	0,03	8,09	6,00	0,04	14,13
64	1	0,03	6,67	6,06	0,07	12,79	1	0,03	5,89	6,02	0,08	11,99

Tamanho dos textos (bytes)	Jaccard						Overlap					
	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,20	394,25	0,87	0,01	395,13	1	0,20	396,03	0,88	0,01	396,93
8192	1	0,10	312,87	1,48	0,02	314,36	1	0,10	313,60	1,45	0,01	315,07
6144	1	0,08	231,10	2,07	0,02	233,18	1	0,08	231,91	1,88	0,02	233,81
4096	1	0,08	156,54	1,86	0,02	158,42	1	0,08	156,66	1,99	0,02	158,66
2048	1	0,06	77,94	2,81	0,12	80,87	1	0,06	79,20	2,79	0,04	82,02
1024	1	0,04	41,15	5,33	0,10	46,58	1	0,04	41,38	5,28	0,05	46,72
512	1	0,04	22,18	5,41	0,11	27,70	1	0,04	22,40	5,42	0,10	27,92
256	1	0,07	12,80	2,08	0,07	14,95	1	0,07	13,07	2,12	0,04	15,23
192	1	0,03	9,95	5,80	0,06	15,80	1	0,03	10,16	5,88	0,03	16,07
128	1	0,03	8,01	5,99	0,04	14,04	1	0,03	8,15	6,04	0,03	14,22
64	1	0,03	5,86	5,99	0,10	11,95	1	0,03	5,98	5,99	0,07	12,04

Tamanho dos textos (bytes)	Simple-matching					
	P	R	Tempo (segundos)			
			Index	Clus	Analis	total
10240	1	0,20	400,24	0,86	0,01	401,11
8192	1	0,10	320,79	1,49	0,02	322,29
6144	1	0,08	237,05	2,02	0,02	239,08
4096	1	0,08	158,43	1,97	0,02	160,42
2048	1	0,06	79,44	2,78	0,03	82,25
1024	1	0,04	40,87	5,94	0,03	46,83
512	1	0,04	22,02	5,39	0,10	27,51
256	1	0,07	12,62	2,10	0,08	14,81
192	1	0,03	10,04	5,78	0,04	15,86
128	1	0,03	7,76	6,01	0,03	13,80
64	1	0,03	6,66	6,00	0,09	12,76

9.8.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	2.484,13	5,95	0,03	2.490,11	1	0,03	2.139,14	4,76	0,04	2.143,94
8192	1	0,03	1.938,28	6,13	0,02	1.944,44	1	0,03	1.673,96	5,43	0,02	1.679,42
6144	1	0,03	1.384,48	6,24	0,02	1.390,74	1	0,03	1.180,16	5,01	0,04	1.185,21
4096	1	0,03	815,81	5,17	0,02	821,00	1	0,03	686,60	5,05	0,02	691,67
2048	1	0,03	350,44	6,09	0,03	356,56	1	0,03	287,03	5,23	0,05	292,31
1024	1	0,03	168,35	5,69	0,03	174,08	1	0,03	137,71	5,34	0,05	143,10
512	1	0,03	83,05	5,35	0,11	88,52	1	0,03	65,06	5,34	0,06	70,46
256	1	0,03	41,60	4,97	0,07	46,65	1	0,03	32,16	4,96	0,11	37,23
192	1	0,03	28,69	5,84	0,05	34,58	1	0,03	21,82	5,88	0,05	27,74
128	1	0,03	19,39	6,04	0,05	25,48	1	0,03	15,36	6,06	0,05	21,47
64	1	0,03	11,08	6,28	0,10	17,45	1	0,03	9,62	6,08	0,09	15,78

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	1.185,80	6,07	0,04	1.191,91	1	0,03	2.163,16	2,85	0,04	2.166,05
8192	1	0,03	840,01	6,18	0,02	846,21	1	0,03	1.685,02	4,39	0,02	1.689,43
6144	1	0,03	506,81	6,32	0,02	513,16	1	0,03	1.192,20	4,02	0,06	1.196,29
4096	1	0,03	278,11	5,21	0,04	283,36	1	0,03	688,86	3,69	0,02	692,57
2048	1	0,03	126,66	5,98	0,04	132,68	1	0,03	286,68	4,22	0,03	290,93
1024	1	0,03	61,98	6,66	0,03	68,67	1	0,03	136,73	5,36	0,05	142,14
512	1	0,03	32,16	5,31	0,09	37,56	1	0,03	65,90	5,33	0,08	71,31
256	1	0,03	17,70	5,01	0,06	22,78	1	0,03	32,11	4,97	0,10	37,18
192	1	0,03	12,10	7,87	0,03	19,99	1	0,03	21,40	5,88	0,03	27,31
128	1	0,03	8,98	6,05	0,02	15,05	1	0,03	15,33	6,04	0,02	21,40
64	1	0,03	5,58	6,09	0,09	11,76	1	0,03	8,62	6,04	0,07	14,72

9.8.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	1	397,23	0,62	0,03	397,88	1	1	394,51	0,65	0,01	395,17
8192	1	1	316,49	0,63	0,01	317,14	1	1	314,88	0,65	0,01	315,55
6144	1	1	231,28	0,63	0,03	231,94	1	1	230,97	0,63	0,01	231,61
4096	1	1	158,87	0,62	0,03	159,52	1	1	156,02	0,61	0,01	156,65
2048	1	0,20	79,84	0,77	0,02	80,63	1	0,20	77,73	0,76	0,03	78,52
1024	1	0,04	42,94	4,57	0,10	47,61	1	0,04	41,18	4,56	0,10	45,84
512	1	0,04	22,99	5,29	0,07	28,34	1	0,04	22,14	5,36	0,09	27,58
256	1	0,07	13,03	2,01	0,06	15,10	1	0,07	12,80	2,05	0,03	14,87
192	1	0,03	10,23	5,65	0,03	15,90	1	0,03	10,01	5,64	0,06	15,71
128	1	0,03	8,25	5,91	0,04	14,20	1	0,03	8,09	5,86	0,04	13,99
64	1	0,03	6,67	5,87	0,07	12,61	1	0,03	5,89	5,88	0,08	11,85

Tamanho dos textos (bytes)	Jaccard						Overlap					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	1	394,25	0,63	0,01	394,90	1	1	396,03	0,63	0,03	396,69
8192	1	1	312,87	0,63	0,01	313,51	1	1	313,60	0,63	0,01	314,24
6144	1	1	231,10	0,64	0,02	231,75	1	1	231,91	0,61	0,01	232,54
4096	1	1	156,54	0,62	0,02	157,18	1	1	156,66	0,61	0,01	157,28
2048	1	0,20	77,94	0,78	0,05	78,76	1	0,20	79,20	0,75	0,03	79,98
1024	1	0,04	41,15	4,52	0,09	45,76	1	0,04	41,38	4,52	0,04	45,94
512	1	0,04	22,18	5,29	0,06	27,53	1	0,04	22,40	5,26	0,08	27,75
256	1	0,07	12,80	2,04	0,03	14,87	1	0,07	13,07	1,99	0,09	15,15
192	1	0,03	9,95	5,66	0,03	15,64	1	0,03	10,16	5,66	0,05	15,87
128	1	0,03	8,01	5,88	0,05	13,95	1	0,03	8,15	5,88	0,06	14,08
64	1	0,03	5,86	5,90	0,08	11,84	1	0,03	5,98	5,86	0,09	11,92

Tamanho dos textos (bytes)	Simple-matching					
			Tempo (segundos)			
	P	R	Index	Clus	Analis	total
10240	1	1	400,24	0,61	0,03	400,88
8192	1	1	320,79	0,61	0,02	321,41
6144	1	1	237,05	0,61	0,01	237,67
4096	1	1	158,43	0,61	0,01	159,06
2048	1	0,20	79,44	0,76	0,03	80,23
1024	1	0,04	40,87	4,54	0,04	45,44
512	1	0,04	22,02	5,23	0,08	27,33
256	1	0,07	12,62	2,02	0,07	14,72
192	1	0,03	10,04	5,67	0,06	15,77
128	1	0,03	7,76	5,90	0,04	13,70
64	1	0,03	6,66	5,93	0,10	12,69

9.8.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	2.484,13	2,24	0,04	2.486,41	1	0,03	2.139,14	4,75	0,03	2.143,92
8192	1	0,03	1.938,28	1,73	0,02	1.940,03	1	0,03	1.673,96	5,36	0,02	1.679,34
6144	1	0,03	1.384,48	1,86	0,04	1.386,38	1	0,03	1.180,16	4,76	0,02	1.184,94
4096	1	0,03	815,81	1,84	0,02	817,67	1	0,03	686,60	5,06	0,04	691,69
2048	1	0,03	350,44	1,53	0,05	352,02	1	0,03	287,03	5,20	0,05	292,28
1024	1	0,03	168,35	1,32	0,05	169,72	1	0,03	137,71	5,31	0,04	143,06
512	1	0,03	83,05	7,05	0,09	90,19	1	0,03	65,06	5,35	0,13	70,54
256	1	0,03	41,60	5,04	0,11	46,75	1	0,03	32,16	4,94	0,08	37,18
192	1	0,03	28,69	1,31	0,03	30,03	1	0,03	21,82	5,87	0,03	27,72
128	1	0,03	19,39	6,18	0,02	25,59	1	0,03	15,36	6,10	0,03	21,49
64	1	0,03	11,08	7,44	0,06	18,57	1	0,03	9,62	6,07	0,08	15,77

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	1.185,80	1,95	0,02	1.187,78	1	0,03	2.163,16	0,77	0,02	2.163,95
8192	1	0,03	840,01	3,41	0,04	843,45	1	0,03	1.685,02	3,53	0,04	1.688,58
6144	1	0,03	506,81	4,70	0,02	511,54	1	0,03	1.192,20	5,89	0,02	1.198,12
4096	1	0,03	278,11	5,14	0,02	283,26	1	0,03	688,86	2,98	0,04	691,87
2048	1	0,03	126,66	6,68	0,05	133,38	1	0,03	286,68	6,07	0,05	292,80
1024	1	0,03	61,98	6,67	0,04	68,70	1	0,03	136,73	6,88	0,02	143,63
512	1	0,03	32,16	5,33	0,06	37,56	1	0,03	65,90	6,17	0,08	72,16
256	1	0,03	17,70	4,95	0,12	22,77	1	0,03	32,11	5,59	0,09	37,79
192	1	0,03	12,10	7,84	0,03	19,97	1	0,03	21,40	5,91	0,04	27,35
128	1	0,03	8,98	6,06	0,02	15,06	1	0,03	15,33	6,05	0,06	21,44
64	1	0,03	5,58	6,02	0,08	11,68	1	0,03	8,62	6,04	0,10	14,75

9.9 APÊNDICE 9: PRIMEIRO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO RSA, COM CHAVES 256 BITS

9.9.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,08	5,98	2,31	0,02	8,31	1	0,08	13,72	2,32	0,02	16,05
8192	1	0,04	11,35	6,71	0,06	18,12	1	0,04	9,00	6,64	0,08	15,71
6144	1	0,04	6,69	6,24	0,07	13,00	1	0,04	5,06	6,26	0,05	11,36
4096	1	0,04	11,29	6,00	0,04	17,33	1	0,04	8,89	5,99	0,07	14,94
2048	1	0,05	10,06	5,00	0,04	15,09	1	0,05	6,04	4,97	0,03	11,04
1024	1	0,05	10,02	4,35	0,02	14,40	1	0,05	7,70	4,42	0,03	12,15
512	1	0,03	10,09	6,54	0,07	16,70	1	0,03	6,87	6,53	0,05	13,45
256	1	0,03	31,85	6,55	0,03	38,42	1	0,03	6,86	6,52	0,02	13,40
192	1	0,03	21,99	6,51	0,02	28,52	1	0,03	6,92	6,56	0,03	13,50
128	1	0,03	10,58	6,63	0,02	17,23	1	0,03	17,13	6,51	0,04	23,69
64	1	0,04	21,97	6,38	0,22	28,57	1	0,04	11,12	6,11	0,03	17,25

	Jaccard						Overlap					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,08	32,08	2,93	0,04	35,04	1	0,08	335,05	2,34	0,02	337,41
8192	1	0,04	31,88	6,62	0,05	38,56	1	0,04	132,63	6,70	0,02	139,35
6144	1	0,04	32,06	6,25	0,08	38,39	1	0,04	336,18	6,23	0,06	342,46
4096	1	0,04	145,47	5,99	0,07	151,52	1	0,04	92,55	6,00	0,05	98,60
2048	1	0,05	103,14	4,94	0,03	108,11	1	0,05	126,63	4,99	0,03	131,65
1024	1	0,05	42,64	4,42	0,02	47,08	1	0,05	124,33	4,38	0,03	128,74
512	1	0,03	102,69	6,56	0,06	109,31	1	0,03	124,00	6,56	0,04	130,60
256	1	0,03	31,37	6,51	0,03	37,91	1	0,03	123,11	6,53	0,04	129,69
192	1	0,03	65,32	6,54	0,04	71,90	1	0,03	631,07	6,53	0,04	637,65
128	1	0,03	62,52	6,53	0,04	69,09	1	0,03	456,90	6,55	0,02	463,47
64	1	0,04	62,32	6,07	0,04	68,44	1	0,04	184,03	7,52	0,02	191,57

	Simple-matching					
Tamanho dos textos (bytes)			Tempo (segundos)			
	P	R	Index	Clus	Analis	total
10240	1	0,08	466,02	2,30	0,02	468,33
8192	1	0,04	122,95	6,71	0,04	129,70
6144	1	0,04	158,98	6,26	0,04	165,29
4096	1	0,04	155,25	5,97	0,06	161,28
2048	1	0,05	157,26	4,97	0,04	162,26
1024	1	0,05	156,17	4,37	0,04	160,58
512	1	0,03	832,59	6,55	0,04	839,17
256	1	0,03	622,95	6,53	0,03	629,50
192	1	0,03	236,77	6,53	0,04	243,34
128	1	0,03	633,81	6,52	0,02	640,35
64	1	0,04	155,85	6,15	0,04	162,04

9.9.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	9,89	7,51	0,02	17,42	1	0,03	62,64	6,34	0,04	69,02
8192	1	0,03	18,76	6,77	0,04	25,57	1	0,03	301,88	6,31	0,03	308,21
6144	1	0,03	17,25	6,30	0,08	23,63	1	0,03	217,79	6,33	0,05	224,17
4096	1	0,03	17,16	6,68	0,06	23,89	1	0,03	89,50	6,31	0,06	95,87
2048	1	0,03	17,51	6,59	0,03	24,13	1	0,03	218,23	6,36	0,05	224,64
1024	1	0,03	69,61	6,33	0,07	76,00	1	0,03	61,62	6,37	0,03	68,02
512	1	0,03	47,98	6,32	0,09	54,38	1	0,03	95,01	6,36	0,03	101,40
256	1	0,03	21,39	6,37	0,03	27,78	1	0,03	91,19	6,34	0,02	97,55
192	1	0,03	48,60	6,35	0,03	54,97	1	0,03	91,91	6,31	0,02	98,24
128	1	0,03	16,90	6,32	0,02	23,24	1	0,03	92,46	6,35	0,03	98,83
64	1	0,03	34,46	6,48	0,04	40,98	1	0,03	462,83	6,41	0,03	469,26

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	6,17	8,13	0,03	14,33	1	0,03	5,27	6,38	0,02	11,67
8192	1	0,03	4,44	7,87	0,04	12,35	1	0,03	5,26	6,34	0,02	11,62
6144	1	0,03	4,44	6,32	0,08	10,84	1	0,03	9,74	6,31	0,04	16,09
4096	1	0,03	4,55	7,64	0,06	12,25	1	0,03	6,39	6,29	0,05	12,73
2048	1	0,03	6,43	8,37	0,04	14,84	1	0,03	3,78	6,33	0,03	10,14
1024	1	0,03	4,68	6,33	0,05	11,05	1	0,03	6,38	6,38	0,03	12,79
512	1	0,03	2,81	6,35	0,04	9,20	1	0,03	5,11	6,35	0,03	11,50
256	1	0,03	4,27	6,35	0,02	10,64	1	0,03	6,52	6,33	0,04	12,89
192	1	0,03	4,32	6,33	0,03	10,67	1	0,03	6,05	6,34	0,04	12,44
128	1	0,03	5,83	6,34	0,04	12,21	1	0,03	6,02	6,34	0,04	12,40
64	1	0,03	5,25	6,34	0,02	11,61	1	0,03	6,06	6,37	0,03	12,46

9.9.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,05	5,98	4,12	0,02	10,12	1	0,05	13,72	4,10	0,02	17,83
8192	1	0,04	11,35	6,57	0,03	17,95	1	0,04	9,00	6,61	0,06	15,67
6144	1	0,04	6,69	6,13	0,07	12,89	1	0,04	5,06	6,15	0,07	11,27
4096	1	0,04	11,29	5,94	0,03	17,25	1	0,04	8,89	5,92	0,07	14,88
2048	1	0,05	10,06	5,25	0,02	15,33	1	0,05	6,04	5,23	0,03	11,30
1024	1	0,05	10,02	4,25	0,05	14,32	1	0,05	7,70	4,26	0,05	12,01
512	1	0,03	10,09	6,41	0,05	16,56	1	0,03	6,87	6,42	0,06	13,35
256	1	0,03	31,85	6,41	0,04	38,30	1	0,03	6,86	6,43	0,04	13,33
192	1	0,03	21,99	6,42	0,04	28,45	1	0,03	6,92	6,44	0,04	13,39
128	1	0,03	10,58	6,42	0,04	17,04	1	0,03	17,13	6,43	0,02	23,58
64	1	0,04	21,97	6,03	0,05	28,05	1	0,04	11,12	6,03	0,05	17,20

Tamanho dos textos (bytes)	Jaccard						Overlap					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,05	32,08	4,10	0,02	36,20	1	0,05	335,05	4,11	0,04	339,20
8192	1	0,04	31,88	6,59	0,03	38,50	1	0,04	132,63	6,58	0,04	139,25
6144	1	0,04	32,06	6,14	0,07	38,27	1	0,04	336,18	6,16	0,07	342,41
4096	1	0,04	145,47	5,94	0,07	151,48	1	0,04	92,55	5,93	0,07	98,54
2048	1	0,05	103,14	5,31	0,02	108,47	1	0,05	126,63	5,24	0,02	131,88
1024	1	0,05	42,64	4,22	0,05	46,91	1	0,05	124,33	4,28	0,05	128,66
512	1	0,03	102,69	6,44	0,04	109,17	1	0,03	124,00	6,42	0,06	130,47
256	1	0,03	31,37	6,42	0,04	37,84	1	0,03	123,11	6,42	0,03	129,56
192	1	0,03	65,32	6,43	0,02	71,78	1	0,03	631,07	6,42	0,02	637,52
128	1	0,03	62,52	6,43	0,03	68,98	1	0,03	456,90	6,42	0,03	463,35
64	1	0,04	62,32	6,04	0,02	68,39	1	0,04	184,03	6,00	0,04	190,07

Tamanho dos textos (bytes)	Simple-matching					
			Tempo (segundos)			
	P	R	Index	Clus	Analis	total
10240	1	0,05	466,02	4,10	0,04	470,15
8192	1	0,04	122,95	6,57	0,02	129,55
6144	1	0,04	158,98	6,13	0,04	165,16
4096	1	0,04	155,25	5,92	0,10	161,27
2048	1	0,05	157,26	5,26	0,04	162,56
1024	1	0,05	156,17	4,27	0,02	160,46
512	1	0,03	832,59	6,41	0,03	839,02
256	1	0,03	622,95	6,41	0,05	629,41
192	1	0,03	236,77	6,42	0,02	243,21
128	1	0,03	633,81	6,42	0,03	640,25
64	1	0,04	155,85	5,94	0,02	161,82

9.9.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	9,89	7,00	0,04	16,93	1	0,03	62,64	5,81	0,02	68,48
8192	1	0,03	18,76	6,66	0,02	25,44	1	0,03	301,88	6,04	0,04	307,96
6144	1	0,03	17,25	5,82	0,05	23,12	1	0,03	217,79	5,85	0,07	223,71
4096	1	0,03	17,16	6,38	0,07	23,61	1	0,03	89,50	6,21	0,06	95,77
2048	1	0,03	17,51	6,03	0,04	23,59	1	0,03	218,23	6,00	0,03	224,26
1024	1	0,03	69,61	5,11	0,03	74,75	1	0,03	61,62	5,11	0,05	66,78
512	1	0,03	47,98	6,21	0,04	54,23	1	0,03	95,01	6,21	0,04	101,26
256	1	0,03	21,39	6,21	0,02	27,62	1	0,03	91,19	6,22	0,03	97,43
192	1	0,03	48,60	6,22	0,03	54,84	1	0,03	91,91	6,22	0,04	98,17
128	1	0,03	16,90	6,21	0,04	23,15	1	0,03	92,46	6,22	0,04	98,72
64	1	0,03	34,46	6,02	0,02	40,51	1	0,03	462,83	6,02	0,03	468,87

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	6,17	7,40	0,04	13,61	1	0,03	5,27	5,88	0,04	11,19
8192	1	0,03	4,44	7,72	0,02	12,18	1	0,03	5,26	6,04	0,04	11,34
6144	1	0,03	4,44	5,85	0,04	10,33	1	0,03	9,74	5,84	0,06	15,64
4096	1	0,03	4,55	7,31	0,07	11,92	1	0,03	6,39	6,29	0,05	12,73
2048	1	0,03	6,43	7,67	0,03	14,12	1	0,03	3,78	6,01	0,04	9,84
1024	1	0,03	4,68	5,11	0,04	9,82	1	0,03	6,38	5,10	0,06	11,53
512	1	0,03	2,81	6,21	0,03	9,05	1	0,03	5,11	6,21	0,04	11,37
256	1	0,03	4,27	6,32	0,04	10,63	1	0,03	6,52	6,22	0,03	12,77
192	1	0,03	4,32	6,20	0,04	10,56	1	0,03	6,05	6,22	0,02	12,30
128	1	0,03	5,83	6,21	0,02	12,07	1	0,03	6,02	6,22	0,02	12,27
64	1	0,03	5,25	6,02	0,04	11,31	1	0,03	6,06	6,02	0,04	12,13

9.9.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,08	5,98	2,02	0,04	8,04	1	0,08	13,72	2,03	0,02	15,76
8192	1	0,04	11,35	6,34	0,05	17,74	1	0,04	9,00	6,41	0,04	15,44
6144	1	0,04	6,69	5,92	0,06	12,68	1	0,04	5,06	5,92	0,04	11,02
4096	1	0,04	11,29	5,68	0,04	17,01	1	0,04	8,89	5,65	0,05	14,59
2048	1	0,05	10,06	4,63	0,04	14,73	1	0,05	6,04	4,64	0,04	10,72
1024	1	0,05	10,02	4,04	0,03	14,09	1	0,05	7,70	4,02	0,03	11,75
512	1	0,03	10,09	6,18	0,06	16,34	1	0,03	6,87	6,18	0,04	13,09
256	1	0,03	31,85	6,21	0,03	38,08	1	0,03	6,86	6,19	0,02	13,08
192	1	0,03	21,99	6,19	0,02	28,20	1	0,03	6,92	6,20	0,02	13,14
128	1	0,03	10,58	6,18	0,02	16,79	1	0,03	17,13	6,20	0,03	23,36
64	1	0,04	21,97	5,78	0,07	27,82	1	0,04	11,12	5,79	0,02	16,94

Tamanho dos textos (bytes)	Jaccard						Overlap					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,08	32,08	2,04	0,02	34,14	1	0,08	335,05	2,01	0,02	337,08
8192	1	0,04	31,88	6,32	0,04	38,24	1	0,04	132,63	6,44	0,03	139,09
6144	1	0,04	32,06	5,94	0,07	38,07	1	0,04	336,18	5,95	0,07	342,20
4096	1	0,04	145,47	5,66	0,07	151,20	1	0,04	92,55	5,65	0,07	98,26
2048	1	0,05	103,14	4,62	0,04	107,80	1	0,05	126,63	4,63	0,04	131,29
1024	1	0,05	42,64	4,05	0,02	46,71	1	0,05	124,33	4,04	0,05	128,42
512	1	0,03	102,69	6,19	0,07	108,95	1	0,03	124,00	6,21	0,05	130,26
256	1	0,03	31,37	6,19	0,02	37,59	1	0,03	123,11	6,20	0,04	129,35
192	1	0,03	65,32	6,19	0,03	71,53	1	0,03	631,07	6,18	0,04	637,30
128	1	0,03	62,52	6,20	0,04	68,76	1	0,03	456,90	6,20	0,04	463,14
64	1	0,04	62,32	5,74	0,04	68,11	1	0,04	184,03	5,77	0,02	189,82

Tamanho dos textos (bytes)	Simple-matching					
			Tempo (segundos)			
	P	R	Index	Clus	Analis	total
10240	1	0,08	466,02	1,99	0,02	468,03
8192	1	0,04	122,95	6,39	0,04	129,39
6144	1	0,04	158,98	5,93	0,03	164,95
4096	1	0,04	155,25	5,66	0,06	160,97
2048	1	0,05	157,26	4,62	0,02	161,90
1024	1	0,05	156,17	4,04	0,04	160,24
512	1	0,03	832,59	6,20	0,04	838,82
256	1	0,03	622,95	6,20	0,07	629,21
192	1	0,03	236,77	6,19	0,04	243,00
128	1	0,03	633,81	6,18	0,04	640,03
64	1	0,04	155,85	5,73	0,04	161,63

9.9.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	9,89	0,90	0,03	10,82	1	0,03	62,64	5,80	0,04	68,48
8192	1	0,03	18,76	1,03	0,04	19,82	1	0,03	301,88	6,03	0,03	307,93
6144	1	0,03	17,25	7,18	0,06	24,49	1	0,03	217,79	5,84	0,05	223,68
4096	1	0,03	17,16	1,49	0,08	18,72	1	0,03	89,50	6,19	0,05	95,74
2048	1	0,03	17,51	0,84	0,02	18,38	1	0,03	218,23	6,01	0,04	224,28
1024	1	0,03	69,61	5,99	0,05	75,65	1	0,03	61,62	5,10	0,03	66,74
512	1	0,03	47,98	7,58	0,07	55,63	1	0,03	95,01	6,20	0,04	101,25
256	1	0,03	21,39	6,48	0,04	27,91	1	0,03	91,19	6,21	0,04	97,44
192	1	0,03	48,60	8,09	0,04	56,72	1	0,03	91,91	6,22	0,02	98,15
128	1	0,03	16,90	7,28	0,02	24,21	1	0,03	92,46	6,21	0,02	98,69
64	1	0,03	34,46	6,46	0,04	40,96	1	0,03	462,83	6,02	0,04	468,89

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	6,17	7,14	0,03	13,33	1	0,03	5,27	7,60	0,02	12,90
8192	1	0,03	4,44	7,70	0,04	12,18	1	0,03	5,26	6,40	0,02	11,68
6144	1	0,03	4,44	5,83	0,06	10,34	1	0,03	9,74	7,21	0,05	16,99
4096	1	0,03	4,55	7,33	0,08	11,96	1	0,03	6,39	7,49	0,09	13,97
2048	1	0,03	6,43	7,67	0,04	14,15	1	0,03	3,78	7,21	0,03	11,02
1024	1	0,03	4,68	5,10	0,04	9,82	1	0,03	6,38	7,79	0,03	14,20
512	1	0,03	2,81	6,22	0,04	9,07	1	0,03	5,11	6,04	0,03	11,18
256	1	0,03	4,27	6,21	0,02	10,50	1	0,03	6,52	6,25	0,04	12,82
192	1	0,03	4,32	6,20	0,02	10,54	1	0,03	6,05	6,22	0,03	12,30
128	1	0,03	5,83	6,20	0,03	12,06	1	0,03	6,02	6,22	0,04	12,29
64	1	0,03	5,25	6,01	0,02	11,28	1	0,03	6,06	6,03	0,02	12,11

9.10 APÊNDICE 10: PRIMEIRO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO RSA, COM CHAVES 512 BITS

9. 10.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	79,24	6,07	0,02	85,33	1	0,03	77,35	6,09	0,04	83,48
8192	1	0,04	63,79	5,35	0,04	69,18	1	0,04	61,25	5,40	0,02	66,67
6144	1	0,03	47,93	6,14	0,04	54,11	1	0,03	45,63	6,10	0,02	51,75
4096	1	0,05	33,91	4,58	0,02	38,51	1	0,05	31,26	4,97	0,02	36,25
2048	1	0,03	18,25	6,80	0,05	25,11	1	0,03	16,93	6,32	0,09	23,34
1024	1	0,03	12,15	6,26	0,10	18,51	1	0,03	10,16	9,59	0,07	19,83
512	1	0,03	7,97	14,31	0,06	22,34	1	0,03	6,95	9,05	0,08	16,08
256	1	0,03	5,86	7,25	0,09	13,21	1	0,03	5,26	7,53	0,10	12,89
192	1	0,03	5,37	6,79	0,07	12,23	1	0,03	4,89	6,25	0,11	11,25
128	1	0,04	4,91	6,18	0,08	11,16	1	0,04	4,47	6,11	0,06	10,64
64	1	0,06	5,43	3,43	0,16	9,01	1	0,06	4,10	3,18	0,04	7,33

	Jaccard						Overlap					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	76,81	6,14	0,02	82,96	1	0,03	76,74	6,07	0,04	82,85
8192	1	0,04	60,70	5,40	0,04	66,14	1	0,04	61,38	5,35	0,02	66,75
6144	1	0,03	45,55	6,06	0,04	51,65	1	0,03	44,97	6,08	0,02	51,07
4096	1	0,05	31,17	4,56	0,02	35,75	1	0,05	31,63	4,62	0,02	36,27
2048	1	0,03	16,80	6,07	0,06	22,93	1	0,03	17,00	6,38	0,07	23,44
1024	1	0,03	10,37	8,99	0,10	19,46	1	0,03	10,22	8,95	0,07	19,23
512	1	0,03	6,92	6,98	0,07	13,97	1	0,03	6,97	11,12	0,09	18,18
256	1	0,03	5,30	7,57	0,06	12,93	1	0,03	5,48	8,08	0,08	13,64
192	1	0,03	4,88	6,74	0,08	11,70	1	0,03	4,92	6,90	0,09	11,91
128	1	0,04	4,51	6,03	0,07	10,60	1	0,04	4,52	6,33	0,06	10,91
64	1	0,06	4,08	3,14	0,06	7,29	1	0,06	4,14	3,18	0,05	7,38

	Simple-matching					
Tamanho dos textos (bytes)	P	R	Tempo (segundos)			
			Index	Clus	Analis	total
10240	1	0,03	76,38	6,23	0,02	82,63
8192	1	0,04	60,48	5,41	0,04	65,93
6144	1	0,03	45,54	6,12	0,02	51,69
4096	1	0,05	31,48	4,36	0,03	35,87
2048	1	0,03	16,53	6,35	0,04	22,92
1024	1	0,03	10,08	7,76	0,08	17,92
512	1	0,03	6,77	7,72	0,11	14,59
256	1	0,03	5,13	11,32	0,08	16,52
192	1	0,03	4,77	6,90	0,07	11,75
128	1	0,04	4,39	6,37	0,09	10,85
64	1	0,06	4,00	3,74	0,07	7,81

9. 10.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

		Euclidiana						Manhattan					
Tamanho dos textos (bytes)	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total

9.10.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	79,24	5,98	0,04	85,25	1	0,03	77,35	5,93	0,02	83,30
8192	1	0,04	63,79	5,21	0,02	69,02	1	0,04	61,25	5,22	0,04	66,51
6144	1	0,03	47,93	5,94	0,02	53,89	1	0,03	45,63	6,02	0,04	51,69
4096	1	0,05	33,91	4,10	0,02	38,03	1	0,05	31,26	4,13	0,02	35,40
2048	1	0,03	18,25	5,95	0,10	24,30	1	0,03	16,93	5,95	0,11	22,98
1024	1	0,03	12,15	6,02	0,05	18,22	1	0,03	10,16	5,94	0,10	16,20
512	1	0,03	7,97	5,97	0,11	14,05	1	0,03	6,95	5,93	0,08	12,96
256	1	0,03	5,86	5,95	0,08	11,89	1	0,03	5,26	5,94	0,06	11,26
192	1	0,03	5,37	5,78	0,11	11,26	1	0,03	4,89	5,80	0,07	10,75
128	1	0,04	4,91	5,34	0,07	10,32	1	0,04	4,47	5,34	0,08	9,88
64	1	0,06	5,43	2,99	0,06	8,47	1	0,06	4,10	2,95	0,07	7,12

Tamanho dos textos (bytes)	Jaccard						Overlap					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	76,81	5,95	0,04	82,80	1	0,03	76,74	5,96	0,02	82,72
8192	1	0,04	60,70	5,23	0,02	65,95	1	0,04	61,38	5,24	0,04	66,66
6144	1	0,03	45,55	6,08	0,02	51,65	1	0,03	44,97	5,96	0,04	50,97
4096	1	0,05	31,17	4,13	0,02	35,32	1	0,05	31,63	4,10	0,03	35,76
2048	1	0,03	16,80	5,97	0,09	22,86	1	0,03	17,00	5,96	0,07	23,03
1024	1	0,03	10,37	5,98	0,07	16,41	1	0,03	10,22	5,92	0,11	16,25
512	1	0,03	6,92	5,92	0,10	12,94	1	0,03	6,97	5,98	0,09	13,04
256	1	0,03	5,30	5,94	0,11	11,35	1	0,03	5,48	5,94	0,08	11,49
192	1	0,03	4,88	5,75	0,07	10,69	1	0,03	4,92	5,89	0,08	10,88
128	1	0,04	4,51	5,34	0,05	9,89	1	0,04	4,52	5,42	0,07	10,02
64	1	0,06	4,08	2,92	0,05	7,06	1	0,06	4,14	2,91	0,07	7,13

Tamanho dos textos (bytes)	Simple-matching					
			Tempo (segundos)			
	P	R	Index	Clus	Analis	total
10240	1	0,03	76,38	5,94	0,02	82,33
8192	1	0,04	60,48	5,22	0,02	65,72
6144	1	0,03	45,54	5,93	0,04	51,51
4096	1	0,05	31,48	4,09	0,02	35,59
2048	1	0,03	16,53	5,94	0,02	22,50
1024	1	0,03	10,08	6,84	0,09	17,01
512	1	0,03	6,77	5,96	0,10	12,83
256	1	0,03	5,13	5,92	0,09	11,13
192	1	0,03	4,77	5,75	0,08	10,61
128	1	0,04	4,39	5,35	0,07	9,81
64	1	0,06	4,00	2,90	0,06	6,95

9. 10.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,03	383,63	6,01	0,04	389,67	1	0,03	275,70	6,00	0,02	281,73
8192	1	0,03	285,96	5,46	0,02	291,44	1	0,03	213,57	5,46	0,12	219,14
6144	1	0,03	222,90	5,98	0,02	228,89	1	0,03	159,40	6,31	0,04	165,75
4096	1	0,03	144,61	5,12	0,02	149,75	1	0,03	100,99	5,11	0,04	106,13
2048	1	0,03	69,07	5,99	0,10	75,16	1	0,03	46,87	6,02	0,07	52,95
1024	1	0,03	33,81	6,00	0,09	39,89	1	0,03	21,62	5,99	0,09	27,70
512	1	0,03	17,71	6,03	0,13	23,88	1	0,03	11,22	6,00	0,09	17,31
256	1	0,03	9,89	5,99	0,08	15,96	1	0,03	6,92	6,00	0,07	12,99
192	1	0,03	8,07	5,97	0,08	14,12	1	0,03	5,32	5,99	0,09	11,40
128	1	0,03	6,65	5,43	0,06	12,14	1	0,03	4,77	5,44	0,07	10,28
64	1	0,03	5,15	5,44	0,06	10,65	1	0,03	3,70	5,45	0,06	9,21

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,03	108,52	6,06	0,04	114,63	1	0,03	275,78	5,99	0,02	281,79
8192	1	0,03	85,14	5,48	0,02	90,64	1	0,03	212,82	5,71	0,04	218,56
6144	1	0,03	63,08	5,99	0,02	69,09	1	0,03	157,48	6,28	0,02	163,78
4096	1	0,03	41,51	5,09	0,02	46,62	1	0,03	101,49	5,21	0,04	106,73
2048	1	0,03	21,13	5,98	0,10	27,21	1	0,03	47,44	6,01	0,04	53,49
1024	1	0,03	10,62	5,97	0,07	16,66	1	0,03	22,69	6,01	0,13	28,83
512	1	0,03	6,22	5,97	0,09	12,27	1	0,03	11,91	5,99	0,08	17,98
256	1	0,03	3,85	6,03	0,06	9,94	1	0,03	6,52	5,99	0,09	12,59
192	1	0,03	3,56	5,99	0,06	9,61	1	0,03	5,32	6,00	0,05	11,37
128	1	0,03	2,95	5,49	0,07	8,51	1	0,03	4,64	5,42	0,04	10,11
64	1	0,03	2,57	5,51	0,05	8,14	1	0,03	4,09	5,41	0,07	9,57

9. 10.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	79,24	5,79	0,02	85,05	1	0,03	77,35	5,82	0,04	83,20
8192	1	0,04	63,79	5,12	0,04	68,95	1	0,04	61,25	5,13	0,02	66,40
6144	1	0,03	47,93	5,84	0,04	53,81	1	0,03	45,63	5,80	0,02	51,45
4096	1	0,05	33,91	3,98	0,04	37,93	1	0,05	31,26	4,00	0,04	35,29
2048	1	0,03	18,25	5,78	0,08	24,11	1	0,03	16,93	5,81	0,09	22,83
1024	1	0,03	12,15	5,82	0,12	18,09	1	0,03	10,16	5,82	0,10	16,08
512	1	0,03	7,97	5,85	0,06	13,87	1	0,03	6,95	5,93	0,08	12,95
256	1	0,03	5,86	5,85	0,06	11,78	1	0,03	5,26	5,82	0,09	11,17
192	1	0,03	5,37	5,60	0,06	11,03	1	0,03	4,89	5,65	0,08	10,61
128	1	0,04	4,91	5,29	0,07	10,27	1	0,04	4,47	5,24	0,06	9,77
64	1	0,06	5,43	2,78	0,05	8,25	1	0,06	4,10	2,79	0,05	6,94

	Jaccard						Overlap					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	76,81	5,82	0,02	82,65	1	0,03	76,74	5,83	0,04	82,61

9. 10.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

Tamanho dos textos (bytes)	Euclidiana						Manhattan					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	383,63	7,71	0,02	391,36	1	0,03	275,70	5,99	0,08	281,77
8192	1	0,03	285,96	8,11	0,04	294,11	1	0,03	213,57	5,43	0,12	219,12
6144	1	0,03	222,90	5,97	0,04	228,90	1	0,03	159,40	5,98	0,02	165,41
4096	1	0,03	144,61	7,54	0,04	152,19	1	0,03	100,99	5,07	0,02	106,08
2048	1	0,03	69,07	7,63	0,06	76,76	1	0,03	46,87	5,95	0,09	52,91
1024	1	0,03	33,81	7,07	0,10	40,98	1	0,03	21,62	5,99	0,07	27,68
512	1	0,03	17,71	6,11	0,08	23,89	1	0,03	11,22	6,01	0,08	17,31
256	1	0,03	9,89	7,54	0,09	17,51	1	0,03	6,92	5,98	0,08	12,98
192	1	0,03	8,07	8,12	0,05	16,24	1	0,03	5,32	5,98	0,06	11,36
128	1	0,03	6,65	6,21	0,09	12,96	1	0,03	4,77	5,45	0,05	10,28
64	1	0,03	5,15	5,41	0,07	10,63	1	0,03	3,70	5,44	0,05	9,20

Tamanho dos textos (bytes)	Canberra						Bray-curtis					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	108,52	5,98	0,03	114,53	1	0,03	275,78	5,95	0,04	281,77
8192	1	0,03	85,14	5,45	0,04	90,62	1	0,03	212,82	6,75	0,02	219,59
6144	1	0,03	63,08	5,98	0,02	69,09	1	0,03	157,48	8,94	0,04	166,46
4096	1	0,03	41,51	5,79	0,04	47,34	1	0,03	101,49	6,48	0,02	107,99
2048	1	0,03	21,13	7,21	0,06	28,40	1	0,03	47,44	6,14	0,02	53,61
1024	1	0,03	10,62	6,00	0,08	16,69	1	0,03	22,69	6,01	0,07	28,76
512	1	0,03	6,22	5,99	0,09	12,30	1	0,03	11,91	5,96	0,14	18,01
256	1	0,03	3,85	5,98	0,11	9,94	1	0,03	6,52	5,95	0,07	12,53
192	1	0,03	3,56	5,98	0,09	9,63	1	0,03	5,32	7,07	0,09	12,48
128	1	0,03	2,95	5,52	0,06	8,53	1	0,03	4,64	5,44	0,08	10,16
64	1	0,03	2,57	5,41	0,08	8,06	1	0,03	4,09	7,98	0,06	12,13

9. 11 APÊNDICE 11: PRIMEIRO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO RSA, COM CHAVES 1024 BITS

9. 11.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice						Co-seno					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	38,17	6,17	0,03	44,36	1	0,03	36,07	6,68	0,02	42,78
8192	1	0,03	31,26	6,23	0,03	37,52	1	0,03	29,25	6,16	0,05	35,46
6144	1	0,03	24,69	6,21	0,04	30,94	1	0,03	22,40	6,77	0,02	29,19
4096	1	0,03	17,84	6,57	0,04	24,45	1	0,03	16,13	6,22	0,03	22,37
2048	1	0,03	11,17	6,72	0,04	17,93	1	0,03	9,69	6,20	0,03	15,92
1024	1	0,03	8,26	6,17	0,02	14,46	1	0,03	6,59	6,41	0,05	13,05
512	1	0,03	6,37	6,15	0,03	12,54	1	0,03	5,01	6,21	0,04	11,26
256	1	0,04	4,69	5,87	0,04	10,60	1	0,04	4,27	6,49	0,02	10,78
192	1	0,03	3,99	6,15	0,03	10,17	1	0,03	3,87	6,12	0,03	10,02
128	1	0,07	3,90	2,96	0,02	6,88	1	0,07	3,94	2,89	0,02	6,86
64	1	0,03	4,06	6,23	0,28	10,57	1	0,03	3,50	6,17	0,05	9,72

	Jaccard						Overlap					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	36,46	6,38	0,04	42,88	1	0,03	36,64	6,19	0,05	42,88
8192	1	0,03	29,23	6,34	0,05	35,63	1	0,03	29,25	6,16	0,05	35,47
6144	1	0,03	22,68	6,17	0,04	28,89	1	0,03	22,69	6,17	0,03	28,88
4096	1	0,03	16,21	6,22	0,04	22,47	1	0,03	16,14	6,21	0,03	22,38
2048	1	0,03	9,68	6,23	0,05	15,96	1	0,03	9,84	6,16	0,03	16,02
1024	1	0,03	6,61	6,13	0,02	12,76	1	0,03	6,71	6,23	0,04	12,98
512	1	0,03	4,99	6,34	0,03	11,36	1	0,03	5,03	6,68	0,04	11,75
256	1	0,04	4,21	5,86	0,04	10,11	1	0,04	4,31	5,83	0,02	10,17
192	1	0,03	4,10	6,15	0,04	10,30	1	0,03	3,90	6,23	0,03	10,16
128	1	0,07	4,00	2,91	0,03	6,94	1	0,07	3,94	2,91	0,02	6,88
64	1	0,03	3,47	6,68	0,07	10,22	1	0,03	3,47	6,24	0,03	9,74

	Simple-matching					
Tamanho dos textos (bytes)			Tempo (segundos)			
	P	R	Index	Clus	Analis	total
10240	1	0,03	36,40	6,17	0,02	42,60
8192	1	0,03	29,08	6,18	0,04	35,30
6144	1	0,03	22,47	6,86	0,04	29,37
4096	1	0,03	15,81	6,19	0,03	22,03
2048	1	0,03	9,55	6,19	0,04	15,78
1024	1	0,03	6,47	6,53	0,04	13,04
512	1	0,03	4,92	6,14	0,04	11,10
256	1	0,04	4,17	5,86	0,04	10,07
192	1	0,03	3,81	6,15	0,03	9,98
128	1	0,07	3,80	2,91	0,02	6,72
64	1	0,03	3,39	6,17	0,03	9,59

9. 11.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Euclidiana			Manhattan		
Tamanho dos			Tempo (segundos)			Tempo (segundos)

9. 11.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	38,17	6,09	0,03	44,28	1	0,03	36,07	6,07	0,04	42,19
8192	1	0,03	31,26	6,08	0,04	37,39	1	0,03	29,25	6,09	0,03	35,36
6144	1	0,03	24,69	6,06	0,03	30,77	1	0,03	22,40	6,06	0,04	28,50
4096	1	0,03	17,84	6,07	0,03	23,93	1	0,03	16,13	6,08	0,05	22,26
2048	1	0,03	11,17	6,08	0,03	17,28	1	0,03	9,69	6,06	0,04	15,80
1024	1	0,03	8,26	6,06	0,04	14,36	1	0,03	6,59	6,05	0,03	12,67
512	1	0,03	6,37	6,07	0,04	12,47	1	0,03	5,01	6,05	0,03	11,08
256	1	0,04	4,69	5,77	0,02	10,48	1	0,04	4,27	5,82	0,04	10,12
192	1	0,03	3,99	6,08	0,02	10,10	1	0,03	3,87	6,08	0,04	9,99
128	1	0,07	3,90	2,81	0,04	6,75	1	0,07	3,94	2,84	0,02	6,80
64	1	0,03	4,06	6,07	0,06	10,20	1	0,03	3,50	6,07	0,09	9,66

Tamanho dos textos (bytes)	Jaccard						Overlap					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	36,46	6,07	0,04	42,56	1	0,03	36,64	6,07	0,02	42,73
8192	1	0,03	29,23	6,06	0,10	35,39	1	0,03	29,25	6,07	0,05	35,36
6144	1	0,03	22,68	6,09	0,03	28,80	1	0,03	22,69	6,07	0,04	28,79
4096	1	0,03	16,21	6,07	0,03	22,31	1	0,03	16,14	6,07	0,02	22,23
2048	1	0,03	9,68	6,07	0,03	15,78	1	0,03	9,84	6,07	0,04	15,95
1024	1	0,03	6,61	6,07	0,04	12,71	1	0,03	6,71	6,07	0,03	12,81
512	1	0,03	4,99	6,07	0,04	11,10	1	0,03	5,03	6,06	0,03	11,12
256	1	0,04	4,21	5,81	0,03	10,05	1	0,04	4,31	5,77	0,04	10,12
192	1	0,03	4,10	6,07	0,03	10,19	1	0,03	3,90	6,07	0,04	10,01
128	1	0,07	4,00	2,84	0,04	6,87	1	0,07	3,94	2,81	0,02	6,77
64	1	0,03	3,47	6,09	0,07	9,63	1	0,03	3,47	6,06	0,05	9,58

Tamanho dos textos (bytes)	Simple-matching					
			Tempo (segundos)			
	P	R	Index	Clus	Analis	total
10240	1	0,03	36,40	6,08	0,04	42,52
8192	1	0,03	29,08	6,06	0,03	35,17
6144	1	0,03	22,47	6,08	0,03	28,57
4096	1	0,03	15,81	6,07	0,04	21,91
2048	1	0,03	9,55	6,07	0,03	15,65
1024	1	0,03	6,47	6,07	0,03	12,57
512	1	0,03	4,92	6,06	0,03	11,02
256	1	0,04	4,17	5,76	0,02	9,95
192	1	0,03	3,81	6,08	0,04	9,93
128	1	0,07	3,80	2,82	0,04	6,65
64	1	0,03	3,39	6,06	0,04	9,49

9. 11.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,03	163,43	5,88	0,04	169,35	1	0,03	110,55	5,89	0,02	116,46
8192	1	0,03	126,45	5,90	0,02	132,37	1	0,03	90,49	5,90	0,04	96,42
6144	1	0,03	94,70	5,88	0,03	100,61	1	0,03	60,90	5,89	0,02	66,82
4096	1	0,03	63,28	5,90	0,04	69,22	1	0,03	42,90	5,90	0,03	48,82
2048	1	0,03	30,97	5,91	0,02	36,91	1	0,03	20,79	5,90	0,05	26,73
1024	1	0,03	16,48	5,89	0,03	22,40	1	0,03	10,87	5,91	0,03	16,80
512	1	0,03	9,39	5,90	0,04	15,33	1	0,03	6,19	5,91	0,04	12,14
256	1	0,03	6,22	5,91	0,02	12,15	1	0,03	4,14	5,88	0,02	10,04
192	1	0,03	4,69	5,90	0,04	10,64	1	0,03	3,05	5,89	0,02	8,97
128	1	0,03	5,15	5,17	0,02	10,34	1	0,03	3,87	5,19	0,05	9,11
64	1	0,03	2,88	5,88	0,04	8,80	1	0,03	2,20	5,91	0,03	8,14

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)	P	R	Tempo (segundos)				P	R	Tempo (segundos)			
			Index	Clus	Analis	total			Index	Clus	Analis	total
10240	1	0,03	50,68	5,89	0,04	56,61	1	0,03	118,51	5,90	0,02	124,44
8192	1	0,03	38,41	5,89	0,04	44,34	1	0,03	90,62	5,88	0,02	96,52
6144	1	0,03	29,79	5,89	0,04	35,71	1	0,03	67,11	5,90	0,05	73,06
4096	1	0,03	19,60	5,90	0,05	25,54	1	0,03	43,47	5,91	0,03	49,40
2048	1	0,03	10,29	5,88	0,03	16,20	1	0,03	20,59	5,89	0,05	26,52
1024	1	0,03	5,83	5,89	0,04	11,76	1	0,03	11,00	5,89	0,04	16,93
512	1	0,03	3,72	5,90	0,03	9,64	1	0,03	6,27	5,89	0,04	12,20
256	1	0,03	2,84	5,89	0,04	8,77	1	0,03	4,13	5,91	0,03	10,07
192	1	0,03	2,25	5,88	0,03	8,16	1	0,03	3,06	5,89	0,03	8,98
128	1	0,03	2,57	5,17	0,03	7,76	1	0,03	3,78	5,18	0,04	9,01
64	1	0,03	1,78	5,89	0,03	7,70	1	0,03	2,24	5,90	0,04	8,18

9. 11.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

Tamanho dos textos (bytes)	Dice						Co-seno					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	38,17	5,88	0,06	44,10	1	0,03	36,07	5,91	0,02	42,01
8192	1	0,03	31,26	5,89	0,03	37,17	1	0,03	29,25	5,88	0,07	35,20
6144	1	0,03	24,69	5,88	0,05	30,62	1	0,03	22,40	5,88	0,02	28,30
4096	1	0,03	17,84	5,90	0,04	23,78	1	0,03	16,13	5,88	0,02	22,02
2048	1	0,03	11,17	5,87	0,04	17,08	1	0,03	9,69	5,90	0,02	15,61
1024	1	0,03	8,26	5,87	0,03	14,15	1	0,03	6,59	5,88	0,04	12,51
512	1	0,03	6,37	5,88	0,03	12,27	1	0,03	5,01	5,87	0,05	10,92
256	1	0,04	4,69	5,62	0,05	10,35	1	0,04	4,27	5,60	0,02	9,88
192	1	0,03	3,99	5,88	0,04	9,91	1	0,03	3,87	5,92	0,03	9,82
128	1	0,07	3,90	2,62	0,02	6,54	1	0,07	3,94	2,62	0,02	6,58
64	1	0,03	4,06	5,92	0,07	10,05	1	0,03	3,50	5,88	0,07	9,45

Tamanho dos textos (bytes)	Jaccard						Overlap					
			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	36,46	5,93	0,02	42,41	1	0,03	36,64	5,87	0,04	42,55
8192	1	0,03	29,23	5,89	0,05	35,17	1	0,03	29,25	5,88	0,08	35,21
6144	1	0,03	22,68	5,87	0,04	28,59	1	0,03	22,69	5,93	0,02	28,64
4096	1	0,03	16,21	5,87	0,04	22,11	1	0,03	16,14	5,88	0,04	22,07
2048	1	0,03	9,68	5,87	0,04	15,59	1	0,03	9,84	5,90	0,03	15,76
1024	1	0,03	6,61	5,90	0,03	12,53	1	0,03	6,71	5,86	0,02	12,60
512	1	0,03	4,99	5,89	0,03	10,90	1	0,03	5,03	5,89	0,04	10,96
256	1	0,04	4,21	5,58	0,04	9,83	1	0,04	4,31	5,60	0,02	9,94
192	1	0,03	4,10	5,87	0,04	10,01	1	0,03	3,90	5,88	0,02	9,81
128	1	0,07	4,00	2,62	0,02	6,64	1	0,07	3,94	2,63	0,02	6,60
64	1	0,03	3,47	5,87	0,07	9,40	1	0,03	3,47	5,88	0,03	9,38

Tamanho dos textos (bytes)	Simple-matching					
			Tempo (segundos)			
	P	R	Index	Clus	Analis	total
10240	1	0,03	36,40	5,86	0,02	42,29
8192	1	0,03	29,08	5,89	0,04	35,01
6144	1	0,03	22,47	5,87	0,04	28,39
4096	1	0,03	15,81	5,89	0,03	21,72
2048	1	0,03	9,55	5,92	0,03	15,50
1024	1	0,03	6,47	5,87	0,02	12,37
512	1	0,03	4,92	5,89	0,04	10,85
256	1	0,04	4,17	5,62	0,04	9,82
192	1	0,03	3,81	5,87	0,03	9,70
128	1	0,07	3,80	2,61	0,02	6,43
64	1	0,03	3,39	5,87	0,02	9,29

9.11.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE DISSIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Euclidiana						Manhattan					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	163,43	5,80	0,03	169,26	1	0,03	110,55	5,91	0,04	116,50
8192	1	0,03	126,45	6,54	0,04	133,02	1	0,03	90,49	5,88	0,02	96,39
6144	1	0,03	94,70	7,26	0,04	102,00	1	0,03	60,90	5,90	0,05	66,85
4096	1	0,03	63,28	7,36	0,04	70,69	1	0,03	42,90	5,89	0,04	48,83
2048	1	0,03	30,97	7,18	0,04	38,19	1	0,03	20,79	5,91	0,03	26,72
1024	1	0,03	16,48	6,10	0,04	22,62	1	0,03	10,87	5,90	0,04	16,81
512	1	0,03	9,39	6,95	0,03	16,36	1	0,03	6,19	5,89	0,03	12,11
256	1	0,03	6,22	6,15	0,04	12,41	1	0,03	4,14	5,89	0,05	10,08
192	1	0,03	4,69	5,89	0,03	10,61	1	0,03	3,05	5,89	0,03	8,97
128	1	0,03	5,15	5,16	0,04	10,35	1	0,03	3,87	5,17	0,03	9,07
64	1	0,03	2,88	7,62	0,04	10,53	1	0,03	2,20	5,90	0,04	8,14

	Canberra						Bray-curtis					
Tamanho dos textos (bytes)			Tempo (segundos)						Tempo (segundos)			
	P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
10240	1	0,03	50,68	5,87	0,02	56,55	1	0,03	118,51	7,40	0,04	125,95
8192	1	0,03	38,41	5,90	0,03	44,31	1	0,03	90,62	6,74	0,02	97,38
6144	1	0,03	29,79	5,90	0,03	35,69	1	0,03	67,11	7,41	0,03	74,55
4096	1	0,03	19,60	5,89	0,03	25,48	1	0,03	43,47	6,70	0,04	50,21
2048	1	0,03	10,29	5,90	0,04	16,19	1	0,03	20,59	5,75	0,02	26,36
1024	1	0,03	5,83	5,89	0,03	11,73	1	0,03	11,00	5,94	0,03	16,97
512	1	0,03	3,72	5,94	0,04	9,65	1	0,03	6,27	5,90	0,03	12,20
256	1	0,03	2,84	5,97	0,03	8,81	1	0,03	4,13	5,89	0,02	10,05
192	1	0,03	2,25	5,90	0,04	8,15	1	0,03	3,06	7,57	0,04	10,67
128	1	0,03	2,57	5,17	0,04	7,73	1	0,03	3,78	6,96	0,02	10,76
64	1	0,03	1,78	5,89	0,04	7,67	1	0,03	2,24	5,90	0,03	8,16

9. 12 APÊNDICE 12: RESULTADO PARA O SEGUNDO CONJUNTO DE EXPERIMENTOS

9.12.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E E DISSIMILARIDADE MÉTODO DE AGRUPAMENTO SINGLE-LINK

Dice				Co-seno				Jaccard			
Tempo (segundos)				Tempo (segundos)				Tempo (segundos)			
Index	Clus	Analis	total	Index	Clus	Analis	total	Index	Clus	Analis	total
3,85	0,13	0,09	4,07	2,89	0,05	0,01	2,95	2,87	0,03	0,01	2,91

Overlap				Simple-matching				Euclidiana			
Tempo (segundos)				Tempo (segundos)				Tempo (segundos)			
Index	Clus	Analis	total	Index	Clus	Analis	total	Index	Clus	Analis	total
2,90	0,03	0,02	2,95	2,93	0,05	0,00	2,98	10,3	0,07	0,009	10,4

Manhattan				Canberra				Bray-Curtis			
Tempo (segundos)				Tempo (segundos)				Tempo (segundos)			
Index	Clus	Analis	total	Index	Clus	Analis	total	Index	Clus	Analis	total
8,11	0,08	0,005	8,2	4,16	0,06	0,01	4,23	8,05	0,07	0,01	8,13

	Dice		Co-seno		Jaccard		Overlap		Simple-matching		Euclidiana		Manhattan		Canberra		Bray-Curtis	
Chave	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R
1	1	1	1	1	1	1	1	1	1	1	0,05	1	1	0,07	0,05	1	1	0,17
2	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	1	0,13	0,05	1	1	0,27
3	1	0,87	1	0,87	1	0,87	1	0,87	1	0,87	0,05	1	1	0,07	0,05	1	1	0,13
4	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	1	0,13	0,05	1	1	0,20
5	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	1	0,07	0,05	1	1	0,27
6	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	1	0,13	0,05	1	1	0,33
7	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	1	0,07	0,05	1	1	0,27
8	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	1	0,13	0,05	1	1	0,27
9	1	0,87	1	0,87	1	0,87	1	0,87	1	0,87	0,05	1	1	0,07	0,05	1	1	0,27
10	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	0,06	1	0,05	1	1	0,40
11	1	0,73	1	0,73	1	0,73	1	0,73	1	0,73	0,05	1	1	0,07	0,05	1	1	0,27
12	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	1	0,13	0,05	1	1	0,20
13	1	0,73	1	0,73	1	0,73	1	0,73	1	0,73	0,05	1	1	0,07	0,05	1	1	0,13
14	1	1	1	1	1	1	1	1	1	1	0,05	1	1	0,13	0,05	1	1	0,20
15	1	0,80	1	0,80	1	0,80	1	0,80	1	0,80	0,05	1	1	0,07	0,05	1	1	0,13
16	1	0,87	1	0,87	1	0,87	1	0,87	1	0,87	0,05	1	1	0,13	0,05	1	1	0,27
17	1	0,80	1	0,80	1	0,80	1	0,80	1	0,80	0,05	1	1	0,07	0,05	1	1	0,20
18	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	1	0,13	0,05	1	1	0,33
19	1	0,73	1	0,73	1	0,73	1	0,73	1	0,73	0,05	1	1	0,07	0,05	1	1	0,20
20	1	0,93	1	0,93	1	0,93	1	0,93	1	0,93	0,05	1	1	0,13	0,05	1	1	0,20

9.12.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E E DISSIMILARIDADE MÉTODO DE AGRUPAMENTO COMPLETE-LINK

Dice				Co-seno				Jaccard			
Tempo (segundos)				Tempo (segundos)				Tempo (segundos)			
Index	Clus	Analis	total	Index	Clus	Analis	total	Index	Clus	Analis	total
3,85	0,08	0,01	3,95	2,89	0,07	0,02	2,98	2,87	0,06	0,01	2,95

Overlap				Simple-matching				Euclidiana			
Tempo (segundos)				Tempo (segundos)				Tempo (segundos)			
Index	Clus	Analis	total	Index	Clus	Analis	total	Index	Clus	Analis	total
2,90	0,06	0,01	2,97	2,93	0,05	0,00	2,97	10,31	0,06	0,01	10,38

Manhattan				Canberra				Bray-Curtis			
Tempo (segundos)				Tempo (segundos)				Tempo (segundos)			
Index	Clus	Analis	total	Index	Clus	Analis	total	Index	Clus	Analis	total
8,11	0,06	0,01	8,18	4,16	0,07	0,01	4,24	8,05	0,04	0,01	8,10

	Dice		Co-seno		Jaccard		Overlap		Simple-matching		Euclidiana		Manhattan		Canberra		Bray-Curtis	
Chave	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R
1	1	0,20	1	0,20	1	0,2	1	0,27	1	0,27	0,1	1	0,1	1	0,05	1	1	0,87
2	1	0,20	1	0,20	1	0,2	1	0,27	1	0,27	0,1	1	0,1	1	0,05	1	1	0,93
3	1	0,13	1	0,13	1	0,1	1	0,20	1	0,20	0,1	1	0,1	1	0,05	1	1	0,67
4	1	0,27	1	0,27	1	0,3	1	0,27	1	0,33	0,1	1	0,1	1	0,05	1	1	0,87
5	1	0,20	1	0,20	1	0,2	1	0,13	1	0,33	0,1	1	0,1	1	0,05	1	1	0,73
6	1	0,40	1	0,40	1	0,4	1	0,27	1	0,40	0,1	1	0,1	1	0,05	1	1	0,87
7	1	0,33	1	0,33	1	0,3	1	0,27	1	0,33	0,1	1	0,1	1	0,05	1	1	0,87
8	1	0,33	1	0,27	1	0,3	1	0,27	1	0,33	0,1	1	0,1	1	0,05	1	1	0,93
9	1	0,27	1	0,20	1	0,3	1	0,13	1	0,27	0,1	1	0,1	1	0,05	1	1	0,73
10	1	0,40	1	0,20	1	0,4	1	0,27	1	0,33	0,1	1	0,1	1	0,05	1	1	0,93
11	1	0,27	1	0,27	1	0,3	1	0,20	1	0,27	0,1	1	0,1	1	0,05	1	1	0,67
12	1	0,40	1	0,27	1	0,4	1	0,27	1	0,40	0,1	1	0,1	1	0,05	1	1	0,80
13	1	0,33	1	0,27	1	0,3	1	0,20	1	0,33	0,1	1	0,1	1	0,05	1	1	0,67
14	1	0,33	1	0,33	1	0,3	1	0,27	1	0,27	0,1	1	0,1	1	0,05	1	1	1,00
15	1	0,20	1	0,20	1	0,2	1	0,20	1	0,27	0,1	1	0,1	1	0,05	1	1	0,73
16	1	0,47	1	0,33	1	0,5	1	0,27	1	0,40	0,1	1	0,1	1	0,05	1	1	0,87
17	1	0,20	1	0,20	1	0,2	1	0,20	1	0,20	0,1	1	0,1	1	0,05	1	1	0,67
18	1	0,33	1	0,27	1	0,3	1	0,27	1	0,33	0,1	1	0,1	1	0,05	1	1	0,93
19	1	0,20	1	0,20	1	0,2	1	0,27	1	0,20	0,1	1	0,1	1	0,05	1	1	0,53
20	1	0,27	1	0,27	1	0,3	1	0,20	1	0,27	0,1	1	0,1	1	0,05	1	1	0,87

9.12.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E E DISSIMILARIDADE MÉTODO DE AGRUPAMENTO GROUP AVERAGE LINK

Dice				Co-seno				Jaccard			
Tempo (segundos)				Tempo (segundos)				Tempo (segundos)			
Index	Clus	Analis	total	Index	Clus	Analis	total	Index	Clus	Analis	total
3,85	0,05	0,01	3,91	2,89	0,05	0,01	2,94	2,87	0,04	0,01	2,92

Overlap				Simple-matching				Euclidiana			
Tempo (segundos)				Tempo (segundos)				Tempo (segundos)			
Index	Clus	Analis	total	Index	Clus	Analis	total	Index	Clus	Analis	total
2,90	0,03	0,00	2,93	2,93	0,04	0,00	2,97	10,31	0,06	0,01	10,38

Manhattan				Canberra				Bray-Curtis			
Tempo (segundos)				Tempo (segundos)				Tempo (segundos)			
Index	Clus	Analis	total	Index	Clus	Analis	total	Index	Clus	Analis	total
8,11	0,06	0,01	8,18	4,16	0,04	0,01	4,21	8,05	0,06	0,01	8,12

	Dice		Co-seno		Jaccard		Overlap		Simple-matching		Euclidiana		Manhattan		Canberra		Bray-Curtis	
Chave	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R
1	1	1	1	1	1	1	1	1	1	1	0,1	1	0,1	1	0,05	1	1	0,40
2	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,47
3	1	0,87	1	0,87	1	0,9	1	0,87	1	0,87	0,1	1	0,1	1	0,05	1	1	0,20
4	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,40
5	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,33
6	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,40
7	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,33
8	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,33
9	1	0,87	1	0,87	1	0,9	1	0,87	1	0,87	0,1	1	0,1	1	0,05	1	1	0,27
10	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,60
11	1	0,73	1	0,73	1	0,7	1	0,73	1	0,73	0,1	1	0,1	1	0,05	1	1	0,27
12	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,53
13	1	0,73	1	0,73	1	0,7	1	0,73	1	0,73	0,1	1	0,1	1	0,05	1	1	0,27
14	1	1	1	1	1	1	1	1	1	1	0,1	1	0,1	1	0,05	1	1	0,53
15	1	0,80	1	0,8	1	0,8	1	0,8	1	0,80	0,1	1	0,1	1	0,05	1	1	0,27
16	1	0,87	1	0,87	1	0,9	1	0,87	1	0,87	0,1	1	0,1	1	0,05	1	1	0,33
17	1	0,80	1	0,8	1	0,8	1	0,8	1	0,80	0,1	1	0,1	1	0,05	1	1	0,20
18	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,33
19	1	0,73	1	0,73	1	0,7	1	0,73	1	0,73	0,1	1	0,1	1	0,05	1	1	0,33
20	1	0,93	1	0,93	1	0,9	1	0,93	1	0,93	0,1	1	0,1	1	0,05	1	1	0,40

9. 13 APÊNDICE 13: RESULTADO PARA O TERCEIRO CONJUNTO DE EXPERIMENTOS

9.13.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E E DISSIMILARIDADE MÉTODO DE AGRUPAMENTO SINGLE-LINK, PARA A CIFRA DE 64 BITS, MODO DE OPERAÇÃO ECB

Dice						Co-seno					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	14,42	0,10	0,09	14,61	1	1	12,81	0,05	0,01	12,87

Jaccard						Overlap					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	12,64	0,03	0,04	12,71	1	1	12,76	0,02	0,00	12,78

Simple-matching					
		Tempo			
P	R	Index	Clus	Analis	total
1	1	12,63	0,02	0,00	12,65

Euclidiana						Manhattan					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,11	1	66,12	0,03	0,01	66,15	1	1	58,29	0,05	0,01	58,34

Canberra						Bray-curtis					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,11	0,75	34,82	0,03	0,01	34,85	1	0,75	57,85	0,02	0,00	57,87

9.13.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E E DISSIMILARIDADE MÉTODO DE AGRUPAMENTO COMPLETE-LINK, PARA A CIFRA DE 64 BITS, MODO DE OPERAÇÃO ECB

Dice						Co-seno					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	14,42	0,03	0,01	14,46	1	1	12,81	0,02	0,01	12,83

Jaccard						Overlap					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	12,64	0,05	0,01	12,70	1	1	12,76	0,02	0,00	12,78

Simple-matching					
		Tempo			
P	R	Index	Clus	Analis	total
1	1	12,63	0,02	0,00	12,65

Euclidiana						Manhattan					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,1	1	66,12	0,02	0,01	66,15	0,1	1	58,29	0,03	0,01	58,32

Canberra						Bray-curtis					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,1	1	34,82	0,04	0,00	34,86	1	1	57,85	0,02	0,00	57,87

9.13.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E E DISSIMILARIDADE MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK, PARA A CIFRA DE 64 BITS, MODO DE OPERAÇÃO ECB

Dice						Co-seno					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	14,42	0,03	0,01	14,46	1	1	12,81	0,02	0,01	12,83

Jaccard						Overlap					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	12,64	0,01	0,00	12,66	1	1	12,76	0,03	0,00	12,80

Simple-matching					
		Tempo			
P	R	Index	Clus	Analis	total
1	1	12,63	0,01	0,00	12,64

Euclidiana						Manhattan					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,1	1	66,12	0,02	0,01	66,15	0,1	1	58,29	0,02	0,01	58,32

Canberra						Bray-curtis					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,1	1	34,82	0,02	0,01	34,84	1	0,75	57,85	0,03	0,00	57,89

9.13.4 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E E DISSIMILARIDADE MÉTODO DE AGRUPAMENTO SINGLE-LINK, PARA A CIFRA DE 128 BITS, MODO DE OPERAÇÃO ECB

Dice						Co-seno					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	7,18	0,10	0,09	7,37	1	1	6,78	0,05	0,02	6,84

Jaccard						Overlap					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	6,81	0,03	0,01	6,84	1	1	6,82	0,03	0,00	6,85

Simple-matching					
		Tempo			
P	R	Index	Clus	Analis	total
1	1	6,70	0,02	0,00	6,72

Euclidiana						Manhattan					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,1	1	34,49	0,03	0,01	34,53	0,11	1	29,54	0,05	0,01	29,59

Canberra						Bray-curtis					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,11	1	15,34	0,03	0,02	15,39	1	0,25	29,80	0,02	0,00	29,82

9.13.5 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E E DISSIMILARIDADE MÉTODO DE AGRUPAMENTO COMPLETE-LINK, PARA A CIFRA DE 128 BITS, MODO DE OPERAÇÃO ECB

Dice						Co-seno					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	0,45	7,18	0,03	0,01	7,22	1	0,45	6,78	0,03	0,01	6,81

Jaccard						Overlap					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	0,45	6,81	0,06	0,01	6,89	1	0,45	6,82	0,02	0,01	6,85

Simple-matching					
		Tempo			
P	R	Index	Clus	Analis	total
1	0,45	6,70	0,02	0,00	6,72

Euclidiana						Manhattan					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,1	1	34,49	0,03	0,01	34,53	0,1	1	29,54	0,03	0,01	29,57

Canberra						Bray-curtis					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,1	1	15,34	0,04	0,00	15,38	1	0,75	29,80	0,02	0,00	29,82

9.13.6 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E E DISSIMILARIDADE MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK, PARA A CIFRA DE 128 BITS, MODO DE OPERAÇÃO ECB

Dice						Co-seno					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	7,18	0,03	0,01	7,21	1	1	6,78	0,02	0,01	6,80

Jaccard						Overlap					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
1	1	6,81	0,01	0,01	6,83	1	1	6,82	0,03	0,00	6,86

Simple-matching					
		Tempo			
P	R	Index	Clus	Analis	total
1	1	6,70	0,01	0,00	6,72

Euclidiana						Manhattan					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,1	1	34,49	0,02	0,01	34,52	0,1	1	29,54	0,02	0,01	29,57

C anberra						Bray-curtis					
		Tempo						Tempo			
P	R	Index	Clus	Analis	total	P	R	Index	Clus	Analis	total
0,1	1	15,34	0,02	0,00	15,37	1	0,4	29,80	0,03	0,00	29,83

9.14 APÊNDICE 14: SEXTO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO AES, COM CHAVES 192 BITS, COM TAMANHO DE TEXTOS MAIORES

9.14.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E DISTÂNCIA, MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice		Co-seno		Jaccard		Overlap		Simple-matching	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R	P	R
20160	1	0,10	1	0,10	1,00	0,10	1	0,10	1	0,10
18592	1	0,50	1	0,50	1,00	0,50	1	0,50	1	0,50
16192	1	0,10	1	0,10	1,00	0,10	1	0,10	1	0,10
14592	1	0,33	1	0,33	1,00	0,33	1	0,33	1	0,33
12240	1	0,07	1	0,07	1,00	0,07	1	0,07	1	0,07

	Euclidiana		Manhattan		Canberra		Bray-curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
20160	0,09	1	1	0,07	0,87	0,13	1	0,03
18592	0,09	1	1	0,10	0,80	0,20	1	0,03
16192	0,05	0,97	1	0,07	0,90	0,10	1	0,03
14592	0,05	1,00	1	0,10	0,54	0,47	1	0,03
12240	0,12	0,93	1	0,07	0,90	0,10	1	0,03

9.14.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E DISTÂNCIA, MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Dice		Co-seno		Jaccard		Overlap		Simple-matching	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R	P	R
20160	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07
18592	1	0,10	1	0,10	1	0,10	1	0,10	1	0,10
16192	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07
14592	1	0,10	1	0,10	1	0,10	1	0,10	1	0,10
12240	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07

	Euclidiana		Manhattan		Canberra		Bray-curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
20160	0,02	1	1	0,10	0,02	1	1	0,03
18592	0,02	1	1	0,50	0,02	1	1	0,03
16192	0,02	1	1	0,10	0,02	1	1	0,03
14592	0,02	1	1	0,33	0,02	1	1	0,03
12240	0,02	1	1	0,07	0,90	0,10	1	0,03

9.14.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E DISTÂNCIAS, MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Dice	Co-seno	Jaccard	Overlap	Simple-matching

9.15 APÊNDICE 15: SEXTO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO AES, COM CHAVES 256 BITS, COM TAMANHO DE TEXTOS MAIORES

9.15.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E DISTÂNCIA, MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice		Co-seno		Jaccard		Overlap		Simple-matching	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R	P	R
20160	1	0,17	1	0,17	1	0,17	1	0,17	1	0,17
18592	1	0,63	1	0,63	1	0,63	1	0,63	1	0,63
16192	1	0,13	1	0,13	1	0,13	1	0,13	1	0,13
14592	1	0,30	1	0,30	1	0,30	1	0,30	1	0,30
12240	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07

	Euclidiana		Manhattan		Canberra		Bray-curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
20160	0,05	1	1	0,07	0,90	0,10	1	0,03
18592	0,09	1	1	0,13	0,90	0,10	1	0,07
16192	0,05	1	1	0,07	0,90	0,10	1	0,03
14592	0,05	1	1	0,10	0,64	0,37	1	0,03
12240	0,05	1	1	0,07	0,84	0,17	1	0,03

9.15.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E DISTÂNCIA, MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Dice		Co-seno		Jaccard		Overlap		Simple-matching	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R	P	R
20160	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07
18592	1	0,13	1	0,13	1	0,13	1	0,13	1	0,13
16192	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07
14592	1	0,10	1	0,10	1	0,10	1	0,10	1	0,10
12240	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07

	Euclidiana		Manhattan		Canberra		Bray-curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
20160	0,02	1	1	0,17	0,02	1	1	0,03
18592	0,02	1	1	0,63	0,02	1	1	0,07
16192	0,02	1	1	0,13	0,02	1	1	0,03
14592	0,02	1	1	0,30	0,02	1	1	0,03
12240	0,02	1	1	0,07	0,02	1	1	0,03

9.15.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Dice		Co-seno		Jaccard		Overlap		Simple-matching	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R	P	R
20160	1	0,17	1	0,17	1	0,17	1	0,17	1	0,17
18592	1	0,63	1	0,63	1	0,63	1	0,63	1	0,63
16192	1	0,13	1	0,13	1	0,13	1	0,13	1	0,13
14592	1	0,30	1	0,30	1	0,30	1	0,30	1	0,30
12240	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07

	Euclidiana		Manhattan		Canberra		Bray-curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
20160	0,02	1	1	0,07	0,81	0,33	1	0,03
18592	0,02	1	1	0,20	0,64	0,50	1	0,07
16192	0,02	1	1	0,07	0,75	1	1	0,03
14592	0,02	1	1	0,10	0,09	1	1	0,03
12240	0,02	1	1	0,07	0,77	0,23	1	0,03

9.16 APÊNDICE 16: SEXTO CONJUNTO DE EXPERIMENTOS: RESULTADO SUBCONJUNTO DE EXPERIMENTOS PARA O ALGORITMO RSA, COM CHAVES 256 BITS, COM TAMANHOS DE TEXTOS MAIORES

9.16.1 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E DISTÂNCIAS, MÉTODO DE AGRUPAMENTO SINGLE-LINK

	Dice		Co-seno		Jaccard		Overlap		Simple-matching	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R	P	R
20160	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07
18600	1	0,63	1	0,63	1	0,63	1	0,63	1	0,63
16200	1	0,13	1	0,13	1	0,13	1	0,13	1	0,13
14600	1	0,67	1	0,67	1	0,67	1	0,67	1	0,67
12240	1	0,13	1	0,13	1	0,13	1	0,13	1	0,13

	Euclidiana		Manhattan		Canberra		Bray-curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
20160	0,02	1	0,02	1	0,02	1	0,02	1
18600	0,02	1	0,02	1	0,02	1	0,02	1
16200	0,02	1	0,02	1	0,02	1	0,02	1
14600	0,02	1	0,02	1	0,02	1	0,02	1
12240	0,02	1	0,02	1	0,02	1	0,02	1

9.16.2 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E DISTÂNCIAS, MÉTODO DE AGRUPAMENTO COMPLETE-LINK

	Dice		Co-seno		Jaccard		Overlap		Simple-matching	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R	P	R
20160	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07
18600	1	0,10	1	0,10	1	0,10	1	0,10	1	0,13
16200	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07
14600	1	0,10	1	0,10	1	0,10	1	0,10	1	0,10
12240	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07

	Euclidiana		Manhattan		Canberra		Bray-curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
20160	1	0,03	0,02	1	1	0,03	0,02	1
18600	1	0,03	0,02	1	1	0,03	0,02	1
16200	1	0,03	0,02	1	1	0,03	0,02	1
14600	1	0,03	0,02	1	1	0,03	0,02	1
12240	1	0,03	0,02	1	1	0,03	0,02	1

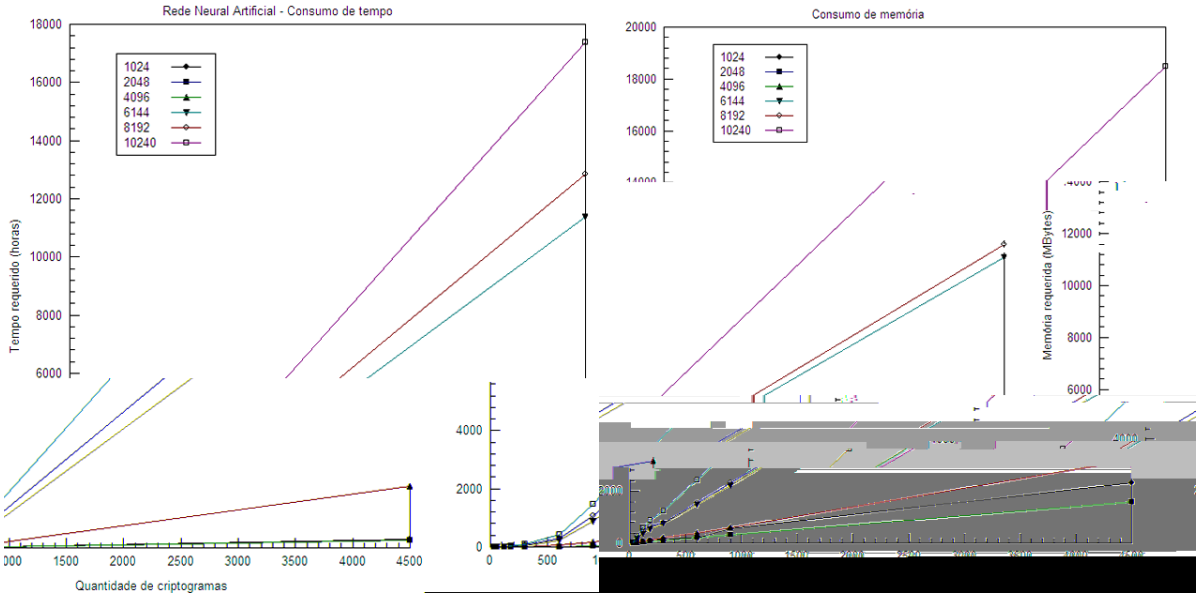
9.16.3 RESULTADO DO AGRUPAMENTO PARA MEDIDAS DE SIMILARIDADE E DISTÂNCIA, MÉTODO DE AGRUPAMENTO GROUP AVERAGE-LINK

	Dice		Co-seno		Jaccard		Overlap		Simple-matching	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R	P	R
20160	1	0,07	1	0,07	1	0,07	1	0,07	1	0,07
18600	1	0,63	1	0,63	1	0,63	1	0,63	1	0,63
16200	1	0,13	1	0,13	1	0,13	1	0,13	1	0,13
14600	1	0,67	1	0,67	1	0,67	1	0,67	1	0,67
12240	1	0,13	1	0,13	1	0,13	1	0,13	1	0,13

	Euclidiana		Manhattan		Canberra		Bray-curtis	
Tamanho dos textos (bytes)	P	R	P	R	P	R	P	R
20160	1	0,03	0,02	1	0,02	1	0,02	1
18600	1	0,03	0,02	1	0,02	1	0,02	1
16200	1	0,03	0,02	1	0,02	1	0,02	1
14600	1	0,03	0,02	1	1	0,03	0,02	1
12240	1	0,03	0,02	1	0,02	1	0,02	1

9.17 APÊNDICE 17: ESTUDO PARCIAL SOBRE O TEMPO DE EXECUÇÃO E CONSUMO DE MEMÓRIA DE UMA REDE NEURAL ARTIFICIAL

Este estudo considerou o sexto conjunto de experimentos. Assim, o tamanho dos criptogramas são os utilizados naquele conjunto. A quantidade de épocas consideradas neste estudo foram cinco.



(a) (b)

GRA. 9.17.1: Tempo de execução e necessidade de memória para a rede neural artificial

Os gráficos 9.17.1, (a) e (b) e as tabelas 9.17.1 a 9.17.6 trazem os resultados dos experimentos preliminares sobre os criptogramas informados. Destes resultados, percebe-se que é inviável a aplicação da rede neural para a tarefa proposta no sexto conjunto de experimentos, uma vez que a quantidade de criptogramas necessários para que o mapa seja corretamente ordenado requerer tempo de execução e memória não disponíveis neste trabalho.

As tabelas 9.17.1 a 9.17.6 trazem em detalhes os tempos e memória requeridos de acordo com o tamanho e a quantidade de criptogramas.

Não é conhecido o número exato de criptogramas e nem com quais tamanhos (em *bytes*) o mapa pode ser ordenado corretamente. Contudo, na figura 9.17.1 pode ser visto o resultado de um experimento realizado sobre os criptogramas aqui considerados, com uma rede 10x10, utilizando 600 criptogramas de 1024 *bytes*.

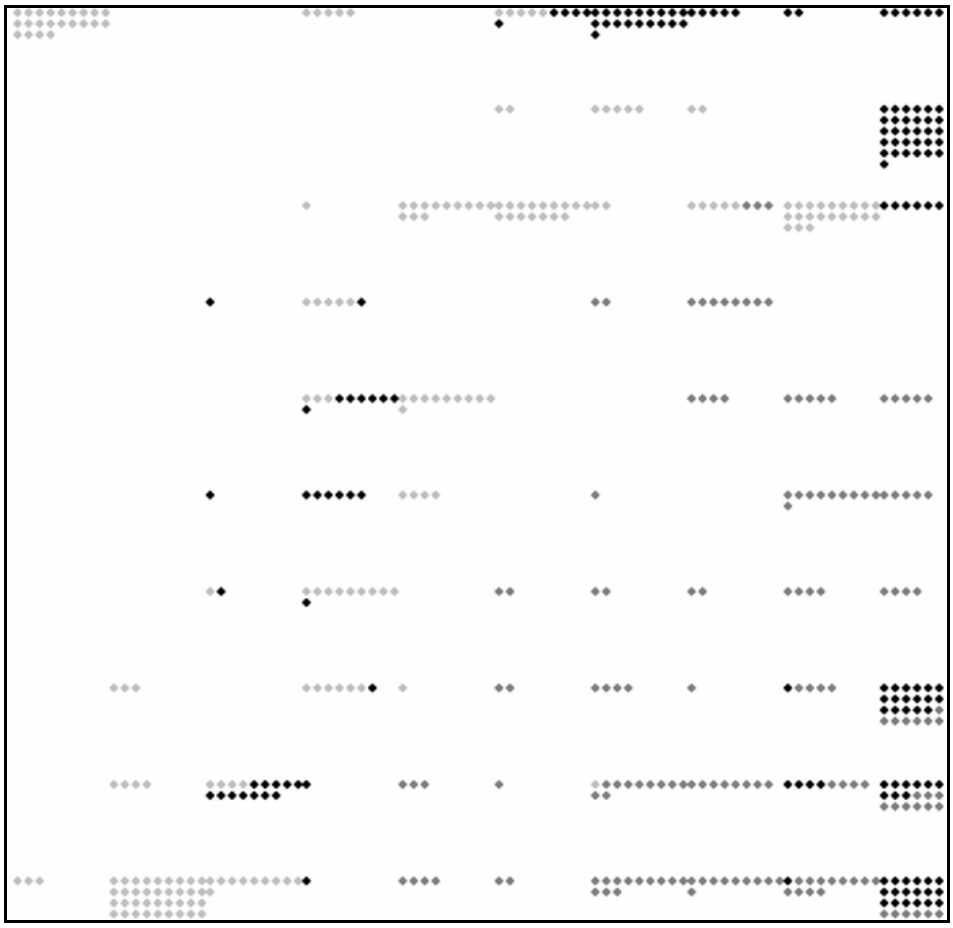


FIG. 9.17.1: Agrupamento de criptogramas cifrados com o AES, DES e RSA

As representações no mapa são feitas por meio de cores, onde os círculos em cinza-claro representando os criptogramas cifrados com o AES, os círculos em preto representando os criptogramas cifrados com o DES e os círculos em cinza representando os criptogramas cifrados com o RSA.

Na figura 9.17.1, pode-se notar que ocorreram poucas imprecisões, significando que poucos neurônios reconheceram mais de um criptograma cifrado com algoritmos diferentes, o que indica um bom potencial de uso da rede neural artificial para agrupamento de padrões em criptogramas, embora, os grupos formados não estão bem definidos ao longo do mapa, ou seja, não ocorreu uma boa ordenação topológica. O esperado era que cada neurônio reconhecesse apenas criptogramas cifrados com um mesmo algoritmo e que os neurônios que reconheceram um mesmo algoritmo ficassem próximos, do posto de vista topológico.

Conclui-se que o mapa de Kohonen tem grande potencial para o agrupamento e classificação de algoritmos criptográficos por meio dos criptogramas gerados por estes

algoritmos. Mas, é necessário aperfeiçoar a rede neural para que a mesma tenha tempo de execução e consumo memória factíveis.

Um estudo feito nesta dissertação sobre o aperfeiçoamento de um módulo para calcular o coeficiente de Spearman, revelou um caminho para a melhoria do desempenho da rede neural. Nesse estudo, para uma base com 1.500 criptogramas, existiam 1.124.250 pares passíveis de possuírem similaridades, considerando que uma medida de similaridade tenha sido usada. Contudo, apenas 1.900 desses pares possuíam similaridade maior que zero. Assim, é necessário aprofundar este estudo para verificar se é conveniente descartar do processo estes pares criptogramas sem similaridades ou o que mais pode ser feito com essa informação.

Abaixo, pode ser visto uma amostra, com os criptogramas obtidos com os 20 primeiros neurônios.

Criptogramas reconhecidos pelo neurônio 11: 22
- Chave obtida: AES-128-3
- Chave obtida: AES-128-4
- Chave obtida: AES-128-4
- Chave obtida: AES-128-4
- Chave obtida: AES-128-5
- Chave obtida: AES-128-5
- Chave obtida: AES-128-5
- Chave obtida: AES-128-5
- Chave obtida: AES-128-6
- Chave obtida: AES-128-6
- Chave obtida: AES-128-6
- Chave obtida: AES-128-6
- Chave obtida: AES-128-7
- Chave obtida: AES-128-7
- Chave obtida: AES-128-7
- Chave obtida: AES-128-8
- Chave obtida: AES-128-8
- Chave obtida: AES-128-8
- Chave obtida: AES-128-9
- Chave obtida: AES-128-9
- Chave obtida: AES-128-9
- Chave obtida: AES-128-9

Criptogramas reconhecidos pelo neurônio 12: 0
Criptogramas reconhecidos pelo neurônio 13: 0
Criptogramas reconhecidos pelo neurônio 14: 5
- Chave obtida: AES-128-7
- Chave obtida: AES-128-8
- Chave obtida: AES-128-8
- Chave obtida: AES-128-9
- Chave obtida: AES-128-9

Criptogramas reconhecidos pelo neurônio 15: 0
Criptogramas reconhecidos pelo neurônio 16: 10
Chave obtida: AES-128-8
Chave obtida: AES-128-8
Chave obtida: AES-128-9
Chave obtida: AES-128-9
Chave obtida: AES-128-9
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-2
Criptogramas reconhecidos pelo neurônio 17: 19
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-1
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Criptogramas reconhecidos pelo neurônio 18: 5
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Criptogramas reconhecidos pelo neurônio 19: 2
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Criptogramas reconhecidos pelo neurônio 110: 9
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Chave obtida: DES-64-2
Chave obtida: DES-64-2

Chave obtida: DES-64-2
Chave obtida: DES-64-2
Criptogramas reconhecidos pelo neurônio 21: 0
Criptogramas reconhecidos pelo neurônio 22: 0
Criptogramas reconhecidos pelo neurônio 23: 0
Criptogramas reconhecidos pelo neurônio 24: 0
Criptogramas reconhecidos pelo neurônio 25: 0
Criptogramas reconhecidos pelo neurônio 26: 2
Chave obtida: AES-128-9
Chave obtida: AES-128-9
Criptogramas reconhecidos pelo neurônio 27: 5
Chave obtida: AES-128-6
Chave obtida: AES-128-7
Chave obtida: AES-128-7
Chave obtida: AES-128-7
Chave obtida: AES-128-8
Criptogramas reconhecidos pelo neurônio 28: 2
Chave obtida: AES-128-7
Chave obtida: AES-128-7
Criptogramas reconhecidos pelo neurônio 29: 0
Criptogramas reconhecidos pelo neurônio 210: 54
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-3
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4

Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-4
Chave obtida: DES-64-5
Chave obtida: DES-64-5
Chave obtida: DES-64-5
Chave obtida: DES-64-5
Chave obtida: DES-64-5
Chave obtida: DES-64-5
Chave obtida: DES-64-5
Chave obtida: DES-64-5
Chave obtida: DES-64-5
Chave obtida: DES-64-5

TAB. 9.17.1: Criptogramas com 1024 *bytes*

Quantidade de Criptogramas	Dimensões Espaço vetorial	Memória requerida (MByte)	Tempo requerido (horas)
4500	565.450	2.340	16.954
900	133.090	551	2.139
600	75.427	228	650
300	37.763	163	158
180	22.618	96	56
120	15.145	65	23
60	7.573	33	6
30	3.836	17	2

TAB. 9.17.2: Criptogramas com 2048 *bytes*

Quantidade de Criptogramas	Dimensões Espaço vetorial	Memória requerida (MByte)	Tempo requerido (horas)
4500	1.100.650	1.568	15.477
900	220.130	314	1.541
600	143.356	204	500
300	73.532	102	160
180	44.026	61	55
120	29.506	42	23
60	14.831	21	4
30	7.493	11	1

TAB. 9.17.3: Criptogramas com 4096 *bytes*

Quantidade de Criptogramas	Dimensões Espaço vetorial	Memória requerida (MByte)	Tempo requerido (horas)
4500	2.272.500	3.192	127.690
900	428.470	602	10.533
600	286.066	402	3.077
300	143.355	201	675
180	85.694	120	176
120	57.643	79	76
60	28.984	41	17
30	14.815	21	3

TAB. 9.17.4: Criptogramas com 6144 *bytes*

Quantidade de Criptogramas	Dimensões Espaço vetorial	Memória requerida (MByte)	Tempo requerido (horas)
4500	3.127.250	11.138	681.974
900	625.450	2.228	52.511
600	417.626	1.487	13.499
300	209.339	746	2.605
180	125.090	533	599
120	84.478	364	249
60	42.759	183	60
30	21.933	31	8

TAB. 9.17.5: Criptogramas com 8192 *bytes*

Quantidade de Criptogramas	Dimensões Espaço vetorial	Memória requerida (MByte)	Tempo requerido (horas)
4500	4.108.900	11.616	771.750
900	821.780	2.323	65.199
600	548.973	1.552	18.398
300	275.445	779	3.899
180	164.356	686	983
120	111.089	480	386
60	55.905	77	68
30	28.963	41	14

TAB. 9.17.6: Criptogramas com 10240 *bytes*

Quantidade de Criptogramas	Dimensões Espaço vetorial	Memória requerida (MByte)	Tempo requerido (horas)
4500	5.152.601	18.494	1.043.012
900	1.014.750	3.642	87.358
600	678.128	2.434	24.828
300	340.398	1.222	5.300
180	202.950	881	1.344
120	137.448	587	637
60	69.279	299	144
30	35.857	51	22

9.18 APÊNDICE 18: CONFIGURAÇÃO DAS MÁQUINAS UTILIZADAS NESTE TRABALHO

9.18.1 COMPUTADOR 1

9.18.1.1 CPU

Especificação: Intel Pentium 4

Velocidade do Clock: 2666.4 MHz

Conjunto de instruções suportadas: MMX, SSE, SSE2

Core: RISC, execução out-of-order e especulativa.

Velocidade do barramento: 533.3 MHz

9.18.1.2 CACHE – L1

Tamanho: 8 KBytes, 4-way set associative, 64 Bytes line size

9.18.1.3 CACHE – L2

Tamanho: 512 KBytes, 8-way set associative, 64 Bytes line size

Velocidade: 2666.4 MHz (Full)

Localização: On Chip

Largura do Barramento: 256 bits

9.18.1.4 MEMÓRIA

Módulo 0: DDR-SDRAM PC3200 - 512 MBytes

Módulo 1: DDR-SDRAM PC3200 - 512 MBytes

9.18.1.5 Placa Mãe

ASUSTeK P4S8X-X, REV 1.xx

9.18.1.6 SISTEMA OPERACIONAL

Windows XP.

9.18.2 COMPUTADOR 2

9.18.2.1 CPU

Especificação: Intel Pentium 4
Velocidade do Clock: 2666.4 MHz
Conjunto de instruções suportadas: MMX, SSE, SSE2
Core: RISC, execução out-of-order e especulativa.
Velocidade do barramento: 533.3 MHz

9.18.2.2 CACHE – L1

Tamanho: 8 KBytes, 4-way set associative, 64 Bytes line size

9.18.2.3 CACHE – L2

Tamanho: 512 KBytes, 8-way set associative, 64 Bytes line size
Velocidade: 2666.4 MHz (Full)
Localização: On Chip
Largura do Barramento: 256 bits

9.18.2.4 MEMÓRIA

Módulo 0: DDR-SDRAM PC3200 - 512 MBytes
Módulo 1: DDR-SDRAM PC3200 - 512 MBytes

9.18.2.5 Placa Mãe

ASUSTeK P4S8X-X, REV 1.xx

9.18.2.6 SISTEMA OPERACIONAL

Windows XP.

9.19 APÊNDICE 19: SÉTIMO CONJUNTO DE EXPERIMENTOS: RESULTADO COM CRIPTOGRAMAS DE 2048 BYTES

9.19.1 FASE DE TREINO (AGRUPAMENTO)

Quantidade de criptogramas obtidos pelo neuronio 11: 12
criptograma obtido: 1: 3_Texto10.txt
criptograma obtido: 2: 3_Texto12.txt
criptograma obtido: 3: 3_Texto15.txt
criptograma obtido: 4: 3_Texto18.txt
criptograma obtido: 5: 3_Texto19.txt
criptograma obtido: 6: 3_Texto20.txt
criptograma obtido: 7: 3_Texto22.txt
criptograma obtido: 8: 3_Texto23.txt
criptograma obtido: 9: 3_Texto4.txt
criptograma obtido: 10: 3_Texto6.txt
criptograma obtido: 11: 3_Texto7.txt
criptograma obtido: 12: 3_Texto9.txt
Quantidade de criptogramas obtidos pelo neuronio 12: 1
criptograma obtido: 1: 3_Texto8.txt
Quantidade de criptogramas obtidos pelo neuronio 13: 2
criptograma obtido: 1: 3_Texto2.txt
criptograma obtido: 2: 3_Texto21.txt
Quantidade de criptogramas obtidos pelo neuronio 14: 2
criptograma obtido: 1: 3_Texto13.txt
criptograma obtido: 2: 3_Texto3.txt
Quantidade de criptogramas obtidos pelo neuronio 15: 1
criptograma obtido: 1: 3_Texto5.txt
Quantidade de criptogramas obtidos pelo neuronio 16: 3
criptograma obtido: 1: 3_Texto14.txt
criptograma obtido: 2: 3_Texto16.txt
criptograma obtido: 3: 3_Texto24.txt
Quantidade de criptogramas obtidos pelo neuronio 17: 3
criptograma obtido: 1: 3_Texto1.txt
criptograma obtido: 2: 3_Texto11.txt
criptograma obtido: 3: 3_Texto17.txt
Quantidade de criptogramas obtidos pelo neuronio 18: 0
Quantidade de criptogramas obtidos pelo neuronio 19: 2
criptograma obtido: 1: 2_Texto4.txt
criptograma obtido: 2: 2_Texto8.txt
Quantidade de criptogramas obtidos pelo neuronio 110: 8
criptograma obtido: 1: 2_Texto10.txt
criptograma obtido: 2: 2_Texto11.txt
criptograma obtido: 3: 2_Texto18.txt
criptograma obtido: 4: 2_Texto19.txt
criptograma obtido: 5: 2_Texto23.txt
criptograma obtido: 6: 2_Texto6.txt
criptograma obtido: 7: 2_Texto7.txt

criptograma obtido: 8: 2_Texto9.txt
Quantidade de criptogramas obtidos pelo neuronio 21: 0
Quantidade de criptogramas obtidos pelo neuronio 22: 0
Quantidade de criptogramas obtidos pelo neuronio 23: 0
Quantidade de criptogramas obtidos pelo neuronio 24: 0
Quantidade de criptogramas obtidos pelo neuronio 25: 0
Quantidade de criptogramas obtidos pelo neuronio 26: 0
Quantidade de criptogramas obtidos pelo neuronio 27: 0
Quantidade de criptogramas obtidos pelo neuronio 28: 0
Quantidade de criptogramas obtidos pelo neuronio 29: 0
Quantidade de criptogramas obtidos pelo neuronio 210: 0
Quantidade de criptogramas obtidos pelo neuronio 31: 0
Quantidade de criptogramas obtidos pelo neuronio 32: 0
Quantidade de criptogramas obtidos pelo neuronio 33: 0
Quantidade de criptogramas obtidos pelo neuronio 34: 0
Quantidade de criptogramas obtidos pelo neuronio 35: 0
Quantidade de criptogramas obtidos pelo neuronio 36: 0
Quantidade de criptogramas obtidos pelo neuronio 37: 0
Quantidade de criptogramas obtidos pelo neuronio 38: 0
Quantidade de criptogramas obtidos pelo neuronio 39: 0
Quantidade de criptogramas obtidos pelo neuronio 310: 0
Quantidade de criptogramas obtidos pelo neuronio 41: 0
Quantidade de criptogramas obtidos pelo neuronio 42: 0
Quantidade de criptogramas obtidos pelo neuronio 43: 0
Quantidade de criptogramas obtidos pelo neuronio 44: 0
Quantidade de criptogramas obtidos pelo neuronio 45: 0
Quantidade de criptogramas obtidos pelo neuronio 46: 0
Quantidade de criptogramas obtidos pelo neuronio 47: 0
Quantidade de criptogramas obtidos pelo neuronio 48: 0
Quantidade de criptogramas obtidos pelo neuronio 49: 2
criptograma obtido: 1: 2_Texto3.txt
criptograma obtido: 2: 2_Texto5.txt
Quantidade de criptogramas obtidos pelo neuronio 410: 1
criptograma obtido: 1: 2_Texto15.txt
Quantidade de criptogramas obtidos pelo neuronio 51: 0
Quantidade de criptogramas obtidos pelo neuronio 52: 0
Quantidade de criptogramas obtidos pelo neuronio 53: 0
Quantidade de criptogramas obtidos pelo neuronio 54: 0
Quantidade de criptogramas obtidos pelo neuronio 55: 0
Quantidade de criptogramas obtidos pelo neuronio 56: 0
Quantidade de criptogramas obtidos pelo neuronio 57: 0
Quantidade de criptogramas obtidos pelo neuronio 58: 0
Quantidade de criptogramas obtidos pelo neuronio 59: 0
Quantidade de criptogramas obtidos pelo neuronio 510: 4
criptograma obtido: 1: 2_Texto2.txt
criptograma obtido: 2: 2_Texto20.txt
criptograma obtido: 3: 2_Texto21.txt
criptograma obtido: 4: 2_Texto24.txt
Quantidade de criptogramas obtidos pelo neuronio 61: 0

Quantidade de criptogramas obtidos pelo neuronio 62: 0
Quantidade de criptogramas obtidos pelo neuronio 63: 0
Quantidade de criptogramas obtidos pelo neuronio 64: 0
Quantidade de criptogramas obtidos pelo neuronio 65: 0
Quantidade de criptogramas obtidos pelo neuronio 66: 0
Quantidade de criptogramas obtidos pelo neuronio 67: 0
Quantidade de criptogramas obtidos pelo neuronio 68: 1
criptograma obtido: 1: 2_Texto22.txt
Quantidade de criptogramas obtidos pelo neuronio 69: 0
Quantidade de criptogramas obtidos pelo neuronio 610: 4
criptograma obtido: 1: 2_Texto13.txt
criptograma obtido: 2: 2_Texto14.txt
criptograma obtido: 3: 2_Texto16.txt
criptograma obtido: 4: 2_Texto17.txt
Quantidade de criptogramas obtidos pelo neuronio 71: 0
Quantidade de criptogramas obtidos pelo neuronio 72: 0
Quantidade de criptogramas obtidos pelo neuronio 73: 0
Quantidade de criptogramas obtidos pelo neuronio 74: 0
Quantidade de criptogramas obtidos pelo neuronio 75: 0
Quantidade de criptogramas obtidos pelo neuronio 76: 0
Quantidade de criptogramas obtidos pelo neuronio 77: 0
Quantidade de criptogramas obtidos pelo neuronio 78: 1
criptograma obtido: 1: 2_Texto12.txt
Quantidade de criptogramas obtidos pelo neuronio 79: 0
Quantidade de criptogramas obtidos pelo neuronio 710: 1
criptograma obtido: 1: 2_Texto1.txt
Quantidade de criptogramas obtidos pelo neuronio 81: 0
Quantidade de criptogramas obtidos pelo neuronio 82: 0
Quantidade de criptogramas obtidos pelo neuronio 83: 0
Quantidade de criptogramas obtidos pelo neuronio 84: 0
Quantidade de criptogramas obtidos pelo neuronio 85: 0
Quantidade de criptogramas obtidos pelo neuronio 86: 0
Quantidade de criptogramas obtidos pelo neuronio 87: 0
Quantidade de criptogramas obtidos pelo neuronio 88: 0
Quantidade de criptogramas obtidos pelo neuronio 89: 0
Quantidade de criptogramas obtidos pelo neuronio 810: 3
criptograma obtido: 1: 1_Texto1.txt
criptograma obtido: 2: 1_Texto10.txt
criptograma obtido: 3: 1_Texto11.txt
Quantidade de criptogramas obtidos pelo neuronio 91: 0
Quantidade de criptogramas obtidos pelo neuronio 92: 0
Quantidade de criptogramas obtidos pelo neuronio 93: 0
Quantidade de criptogramas obtidos pelo neuronio 94: 0
Quantidade de criptogramas obtidos pelo neuronio 95: 0
Quantidade de criptogramas obtidos pelo neuronio 96: 0
Quantidade de criptogramas obtidos pelo neuronio 97: 0
Quantidade de criptogramas obtidos pelo neuronio 98: 0
Quantidade de criptogramas obtidos pelo neuronio 99: 2
criptograma obtido: 1: 1_Texto12.txt

criptograma obtido: 2: 1_Texto13.txt
Quantidade de criptogramas obtidos pelo neuronio 910: 3
criptograma obtido: 1: 1_Texto16.txt
criptograma obtido: 2: 1_Texto18.txt
criptograma obtido: 3: 1_Texto20.txt
Quantidade de criptogramas obtidos pelo neuronio 101: 0
Quantidade de criptogramas obtidos pelo neuronio 102: 0
Quantidade de criptogramas obtidos pelo neuronio 103: 0
Quantidade de criptogramas obtidos pelo neuronio 104: 0
Quantidade de criptogramas obtidos pelo neuronio 105: 0
Quantidade de criptogramas obtidos pelo neuronio 106: 0
Quantidade de criptogramas obtidos pelo neuronio 107: 6
criptograma obtido: 1: 1_Texto14.txt
criptograma obtido: 2: 1_Texto15.txt
criptograma obtido: 3: 1_Texto17.txt
criptograma obtido: 4: 1_Texto19.txt
criptograma obtido: 5: 1_Texto23.txt
criptograma obtido: 6: 5_Texto1.txt
Quantidade de criptogramas obtidos pelo neuronio 108: 6
criptograma obtido: 1: 1_Texto2.txt
criptograma obtido: 2: 1_Texto21.txt
criptograma obtido: 3: 1_Texto22.txt
criptograma obtido: 4: 4_Texto1.txt
criptograma obtido: 5: 5_Texto11.txt
criptograma obtido: 6: 5_Texto12.txt
Quantidade de criptogramas obtidos pelo neuronio 109: 5
criptograma obtido: 1: 1_Texto24.txt
criptograma obtido: 2: 1_Texto6.txt
criptograma obtido: 3: 4_Texto10.txt
criptograma obtido: 4: 4_Texto11.txt
criptograma obtido: 5: 4_Texto12.txt
Quantidade de criptogramas obtidos pelo neuronio 1010: 47
criptograma obtido: 1: 1_Texto3.txt
criptograma obtido: 2: 1_Texto4.txt
criptograma obtido: 3: 1_Texto5.txt
criptograma obtido: 4: 1_Texto7.txt
criptograma obtido: 5: 1_Texto8.txt
criptograma obtido: 6: 1_Texto9.txt
criptograma obtido: 7: 4_Texto13.txt
criptograma obtido: 8: 4_Texto14.txt
criptograma obtido: 9: 4_Texto15.txt
criptograma obtido: 10: 4_Texto16.txt
criptograma obtido: 11: 4_Texto17.txt
criptograma obtido: 12: 4_Texto18.txt
criptograma obtido: 13: 4_Texto19.txt
criptograma obtido: 14: 4_Texto2.txt
criptograma obtido: 15: 4_Texto20.txt
criptograma obtido: 16: 4_Texto21.txt
criptograma obtido: 17: 4_Texto22.txt

criptograma obtido: 18: 4_Texto23.txt
criptograma obtido: 19: 4_Texto24.txt
criptograma obtido: 20: 4_Texto3.txt
criptograma obtido: 21: 4_Texto4.txt
criptograma obtido: 22: 4_Texto5.txt
criptograma obtido: 23: 4_Texto6.txt
criptograma obtido: 24: 4_Texto7.txt
criptograma obtido: 25: 4_Texto8.txt
criptograma obtido: 26: 4_Texto9.txt
criptograma obtido: 27: 5_Texto10.txt
criptograma obtido: 28: 5_Texto13.txt
criptograma obtido: 29: 5_Texto14.txt
criptograma obtido: 30: 5_Texto15.txt
criptograma obtido: 31: 5_Texto16.txt
criptograma obtido: 32: 5_Texto17.txt
criptograma obtido: 33: 5_Texto18.txt
criptograma obtido: 34: 5_Texto19.txt
criptograma obtido: 35: 5_Texto2.txt
criptograma obtido: 36: 5_Texto20.txt
criptograma obtido: 37: 5_Texto21.txt
criptograma obtido: 38: 5_Texto22.txt
criptograma obtido: 39: 5_Texto23.txt
criptograma obtido: 40: 5_Texto24.txt
criptograma obtido: 41: 5_Texto3.txt
criptograma obtido: 42: 5_Texto4.txt
criptograma obtido: 43: 5_Texto5.txt
criptograma obtido: 44: 5_Texto6.txt
criptograma obtido: 45: 5_Texto7.txt
criptograma obtido: 46: 5_Texto8.txt

9.19.2 FASE DE TESTE (CLASSIFICAÇÃO)

Quantidade de criptogramas obtidos pelo neuronio 11: 0
Quantidade de criptogramas obtidos pelo neuronio 12: 1
criptograma obtido: 1: 3_Texto25.txt
Quantidade de criptogramas obtidos pelo neuronio 13: 1
criptograma obtido: 1: 3_Texto27.txt
Quantidade de criptogramas obtidos pelo neuronio 14: 4
criptograma obtido: 1: 3_Texto26.txt
criptograma obtido: 2: 3_Texto28.txt
criptograma obtido: 3: 3_Texto29.txt
criptograma obtido: 4: 3_Texto30.txt
Quantidade de criptogramas obtidos pelo neuronio 15: 0
Quantidade de criptogramas obtidos pelo neuronio 16: 0
Quantidade de criptogramas obtidos pelo neuronio 17: 0
Quantidade de criptogramas obtidos pelo neuronio 18: 0
Quantidade de criptogramas obtidos pelo neuronio 19: 0
Quantidade de criptogramas obtidos pelo neuronio 110: 0

Quantidade de criptogramas obtidos pelo neuronio 21: 0
Quantidade de criptogramas obtidos pelo neuronio 22: 0
Quantidade de criptogramas obtidos pelo neuronio 23: 0
Quantidade de criptogramas obtidos pelo neuronio 24: 0
Quantidade de criptogramas obtidos pelo neuronio 25: 0
Quantidade de criptogramas obtidos pelo neuronio 26: 0
Quantidade de criptogramas obtidos pelo neuronio 27: 0
Quantidade de criptogramas obtidos pelo neuronio 28: 0
Quantidade de criptogramas obtidos pelo neuronio 29: 0
Quantidade de criptogramas obtidos pelo neuronio 210: 0
Quantidade de criptogramas obtidos pelo neuronio 31: 0
Quantidade de criptogramas obtidos pelo neuronio 32: 0
Quantidade de criptogramas obtidos pelo neuronio 33: 0
Quantidade de criptogramas obtidos pelo neuronio 34: 0
Quantidade de criptogramas obtidos pelo neuronio 35: 0
Quantidade de criptogramas obtidos pelo neuronio 36: 0
Quantidade de criptogramas obtidos pelo neuronio 37: 0
Quantidade de criptogramas obtidos pelo neuronio 38: 0
Quantidade de criptogramas obtidos pelo neuronio 39: 0
Quantidade de criptogramas obtidos pelo neuronio 310: 0
Quantidade de criptogramas obtidos pelo neuronio 41: 0
Quantidade de criptogramas obtidos pelo neuronio 42: 0
Quantidade de criptogramas obtidos pelo neuronio 43: 0
Quantidade de criptogramas obtidos pelo neuronio 44: 0
Quantidade de criptogramas obtidos pelo neuronio 45: 0
Quantidade de criptogramas obtidos pelo neuronio 46: 0
Quantidade de criptogramas obtidos pelo neuronio 47: 0
Quantidade de criptogramas obtidos pelo neuronio 48: 0
Quantidade de criptogramas obtidos pelo neuronio 49: 0
Quantidade de criptogramas obtidos pelo neuronio 410: 0
Quantidade de criptogramas obtidos pelo neuronio 51: 0
Quantidade de criptogramas obtidos pelo neuronio 52: 0
Quantidade de criptogramas obtidos pelo neuronio 53: 0
Quantidade de criptogramas obtidos pelo neuronio 54: 0
Quantidade de criptogramas obtidos pelo neuronio 55: 0
Quantidade de criptogramas obtidos pelo neuronio 56: 0
Quantidade de criptogramas obtidos pelo neuronio 57: 0
Quantidade de criptogramas obtidos pelo neuronio 58: 0
Quantidade de criptogramas obtidos pelo neuronio 59: 0
Quantidade de criptogramas obtidos pelo neuronio 510: 6
    criptograma obtido: 1: 2_Texto25.txt
    criptograma obtido: 2: 2_Texto26.txt
    criptograma obtido: 3: 2_Texto27.txt
    criptograma obtido: 4: 2_Texto28.txt
    criptograma obtido: 5: 2_Texto29.txt
    criptograma obtido: 6: 2_Texto30.txt
Quantidade de criptogramas obtidos pelo neuronio 61: 0
Quantidade de criptogramas obtidos pelo neuronio 62: 0
Quantidade de criptogramas obtidos pelo neuronio 63: 0

Quantidade de criptogramas obtidos pelo neuronio 64: 0
Quantidade de criptogramas obtidos pelo neuronio 65: 0
Quantidade de criptogramas obtidos pelo neuronio 66: 0
Quantidade de criptogramas obtidos pelo neuronio 67: 0
Quantidade de criptogramas obtidos pelo neuronio 68: 0
Quantidade de criptogramas obtidos pelo neuronio 69: 0
Quantidade de criptogramas obtidos pelo neuronio 610: 0
Quantidade de criptogramas obtidos pelo neuronio 71: 0
Quantidade de criptogramas obtidos pelo neuronio 72: 0
Quantidade de criptogramas obtidos pelo neuronio 73: 0
Quantidade de criptogramas obtidos pelo neuronio 74: 0
Quantidade de criptogramas obtidos pelo neuronio 75: 0
Quantidade de criptogramas obtidos pelo neuronio 76: 0
Quantidade de criptogramas obtidos pelo neuronio 77: 0
Quantidade de criptogramas obtidos pelo neuronio 78: 0
Quantidade de criptogramas obtidos pelo neuronio 79: 0
Quantidade de criptogramas obtidos pelo neuronio 710: 0
Quantidade de criptogramas obtidos pelo neuronio 81: 0
Quantidade de criptogramas obtidos pelo neuronio 82: 0
Quantidade de criptogramas obtidos pelo neuronio 83: 0
Quantidade de criptogramas obtidos pelo neuronio 84: 0
Quantidade de criptogramas obtidos pelo neuronio 85: 0
Quantidade de criptogramas obtidos pelo neuronio 86: 0
Quantidade de criptogramas obtidos pelo neuronio 87: 0
Quantidade de criptogramas obtidos pelo neuronio 88: 0
Quantidade de criptogramas obtidos pelo neuronio 89: 0
Quantidade de criptogramas obtidos pelo neuronio 810: 0
Quantidade de criptogramas obtidos pelo neuronio 91: 0
Quantidade de criptogramas obtidos pelo neuronio 92: 0
Quantidade de criptogramas obtidos pelo neuronio 93: 0
Quantidade de criptogramas obtidos pelo neuronio 94: 0
Quantidade de criptogramas obtidos pelo neuronio 95: 0
Quantidade de criptogramas obtidos pelo neuronio 96: 0
Quantidade de criptogramas obtidos pelo neuronio 97: 0
Quantidade de criptogramas obtidos pelo neuronio 98: 0
Quantidade de criptogramas obtidos pelo neuronio 99: 0
Quantidade de criptogramas obtidos pelo neuronio 910: 0
Quantidade de criptogramas obtidos pelo neuronio 101: 0
Quantidade de criptogramas obtidos pelo neuronio 102: 0
Quantidade de criptogramas obtidos pelo neuronio 103: 0
Quantidade de criptogramas obtidos pelo neuronio 104: 0
Quantidade de criptogramas obtidos pelo neuronio 105: 0
Quantidade de criptogramas obtidos pelo neuronio 106: 0
Quantidade de criptogramas obtidos pelo neuronio 107: 0
Quantidade de criptogramas obtidos pelo neuronio 108: 6
criptograma obtido: 1: 1_Texto25.txt
criptograma obtido: 2: 1_Texto26.txt
criptograma obtido: 3: 1_Texto27.txt
criptograma obtido: 4: 1_Texto28.txt

criptograma obtido: 5: 1_Texto29.txt
criptograma obtido: 6: 1_Texto30.txt
Quantidade de criptogramas obtidos pelo neuronio 109: 0
Quantidade de criptogramas obtidos pelo neuronio 1010: 12
criptograma obtido: 1: 4_Texto25.txt
criptograma obtido: 2: 4_Texto26.txt
criptograma obtido: 3: 4_Texto27.txt
criptograma obtido: 4: 4_Texto28.txt
criptograma obtido: 5: 4_Texto29.txt
criptograma obtido: 6: 4_Texto30.txt
criptograma obtido: 7: 5_Texto25.txt
criptograma obtido: 8: 5_Texto26.txt
criptograma obtido: 9: 5_Texto27.txt
criptograma obtido: 10: 5_Texto28.txt
criptograma obtido: 11: 5_Texto29.txt
criptograma obtido: 12: 5_Texto30.txt

9.20 APÊNDICE 20: SÉTIMO CONJUNTO DE EXPERIMENTOS: RESULTADO COM CRIPTOGRAMAS DE 6144 BYTES

9.20.1 FASE DE TREINO (AGRUPAMENTO)

Quantidade de criptogramas obtidos pelo neuronio 11: 13
- criptograma obtido: 1: 8_Texto10.txt
- criptograma obtido: 2: 8_Texto11.txt
- criptograma obtido: 3: 8_Texto13.txt
- criptograma obtido: 4: 8_Texto14.txt
- criptograma obtido: 5: 8_Texto18.txt
- criptograma obtido: 6: 8_Texto2.txt
- criptograma obtido: 7: 8_Texto21.txt
- criptograma obtido: 8: 8_Texto22.txt
- criptograma obtido: 9: 8_Texto23.txt
- criptograma obtido: 10: 8_Texto24.txt
- criptograma obtido: 11: 8_Texto5.txt
- criptograma obtido: 12: 8_Texto8.txt
- criptograma obtido: 13: 8_Texto9.txt

Quantidade de criptogramas obtidos pelo neuronio 12: 2
- criptograma obtido: 1: 8_Texto6.txt
- criptograma obtido: 2: 8_Texto7.txt

Quantidade de criptogramas obtidos pelo neuronio 13: 2
- criptograma obtido: 1: 8_Texto16.txt
- criptograma obtido: 2: 8_Texto17.txt

Quantidade de criptogramas obtidos pelo neuronio 14: 0

Quantidade de criptogramas obtidos pelo neuronio 15: 7
- criptograma obtido: 1: 8_Texto1.txt
- criptograma obtido: 2: 8_Texto12.txt
- criptograma obtido: 3: 8_Texto15.txt
- criptograma obtido: 4: 8_Texto19.txt
- criptograma obtido: 5: 8_Texto20.txt
- criptograma obtido: 6: 8_Texto3.txt
- criptograma obtido: 7: 8_Texto4.txt

Quantidade de criptogramas obtidos pelo neuronio 16: 0

Quantidade de criptogramas obtidos pelo neuronio 17: 0

Quantidade de criptogramas obtidos pelo neuronio 18: 0

Quantidade de criptogramas obtidos pelo neuronio 19: 0

Quantidade de criptogramas obtidos pelo neuronio 110: 13
- criptograma obtido: 1: 5_Texto11.txt
- criptograma obtido: 2: 5_Texto14.txt
- criptograma obtido: 3: 5_Texto19.txt
- criptograma obtido: 4: 5_Texto2.txt
- criptograma obtido: 5: 5_Texto20.txt
- criptograma obtido: 6: 5_Texto21.txt
- criptograma obtido: 7: 5_Texto22.txt
- criptograma obtido: 8: 5_Texto23.txt
- criptograma obtido: 9: 5_Texto24.txt

criptograma obtido: 10: 5_Texto5.txt
criptograma obtido: 11: 5_Texto6.txt
criptograma obtido: 12: 5_Texto8.txt
criptograma obtido: 13: 5_Texto9.txt
Quantidade de criptogramas obtidos pelo neuronio 21: 0
Quantidade de criptogramas obtidos pelo neuronio 22: 0
Quantidade de criptogramas obtidos pelo neuronio 23: 0
Quantidade de criptogramas obtidos pelo neuronio 24: 0
Quantidade de criptogramas obtidos pelo neuronio 25: 0
Quantidade de criptogramas obtidos pelo neuronio 26: 0
Quantidade de criptogramas obtidos pelo neuronio 27: 0
Quantidade de criptogramas obtidos pelo neuronio 28: 0
Quantidade de criptogramas obtidos pelo neuronio 29: 0
Quantidade de criptogramas obtidos pelo neuronio 210: 3
criptograma obtido: 1: 5_Texto10.txt
criptograma obtido: 2: 5_Texto18.txt
criptograma obtido: 3: 5_Texto4.txt
Quantidade de criptogramas obtidos pelo neuronio 31: 0
Quantidade de criptogramas obtidos pelo neuronio 32: 0
Quantidade de criptogramas obtidos pelo neuronio 33: 0
Quantidade de criptogramas obtidos pelo neuronio 34: 0
Quantidade de criptogramas obtidos pelo neuronio 35: 0
Quantidade de criptogramas obtidos pelo neuronio 36: 0
Quantidade de criptogramas obtidos pelo neuronio 37: 0
Quantidade de criptogramas obtidos pelo neuronio 38: 0
Quantidade de criptogramas obtidos pelo neuronio 39: 0
Quantidade de criptogramas obtidos pelo neuronio 310: 8
criptograma obtido: 1: 5_Texto1.txt
criptograma obtido: 2: 5_Texto12.txt
criptograma obtido: 3: 5_Texto13.txt
criptograma obtido: 4: 5_Texto15.txt
criptograma obtido: 5: 5_Texto16.txt
criptograma obtido: 6: 5_Texto17.txt
criptograma obtido: 7: 5_Texto3.txt
criptograma obtido: 8: 5_Texto7.txt
Quantidade de criptogramas obtidos pelo neuronio 41: 0
Quantidade de criptogramas obtidos pelo neuronio 42: 0
Quantidade de criptogramas obtidos pelo neuronio 43: 0
Quantidade de criptogramas obtidos pelo neuronio 44: 0
Quantidade de criptogramas obtidos pelo neuronio 45: 0
Quantidade de criptogramas obtidos pelo neuronio 46: 0
Quantidade de criptogramas obtidos pelo neuronio 47: 0
Quantidade de criptogramas obtidos pelo neuronio 48: 0
Quantidade de criptogramas obtidos pelo neuronio 49: 0
Quantidade de criptogramas obtidos pelo neuronio 410: 0
Quantidade de criptogramas obtidos pelo neuronio 51: 0
Quantidade de criptogramas obtidos pelo neuronio 52: 0
Quantidade de criptogramas obtidos pelo neuronio 53: 0
Quantidade de criptogramas obtidos pelo neuronio 54: 0

Quantidade de criptogramas obtidos pelo neuronio 55: 0
Quantidade de criptogramas obtidos pelo neuronio 56: 0
Quantidade de criptogramas obtidos pelo neuronio 57: 0
Quantidade de criptogramas obtidos pelo neuronio 58: 0
Quantidade de criptogramas obtidos pelo neuronio 59: 0
Quantidade de criptogramas obtidos pelo neuronio 510: 0
Quantidade de criptogramas obtidos pelo neuronio 61: 0
Quantidade de criptogramas obtidos pelo neuronio 62: 0
Quantidade de criptogramas obtidos pelo neuronio 63: 0
Quantidade de criptogramas obtidos pelo neuronio 64: 0
Quantidade de criptogramas obtidos pelo neuronio 65: 0
Quantidade de criptogramas obtidos pelo neuronio 66: 0
Quantidade de criptogramas obtidos pelo neuronio 67: 0
Quantidade de criptogramas obtidos pelo neuronio 68: 0
Quantidade de criptogramas obtidos pelo neuronio 69: 0
Quantidade de criptogramas obtidos pelo neuronio 610: 24
criptograma obtido: 1: 3_Texto1.txt
criptograma obtido: 2: 3_Texto10.txt
criptograma obtido: 3: 3_Texto11.txt
criptograma obtido: 4: 3_Texto12.txt
criptograma obtido: 5: 3_Texto13.txt
criptograma obtido: 6: 3_Texto14.txt
criptograma obtido: 7: 3_Texto15.txt
criptograma obtido: 8: 3_Texto16.txt
criptograma obtido: 9: 3_Texto17.txt
criptograma obtido: 10: 3_Texto18.txt
criptograma obtido: 11: 3_Texto19.txt
criptograma obtido: 12: 3_Texto2.txt
criptograma obtido: 13: 3_Texto20.txt
criptograma obtido: 14: 3_Texto21.txt
criptograma obtido: 15: 3_Texto22.txt
criptograma obtido: 16: 3_Texto23.txt
criptograma obtido: 17: 3_Texto24.txt
criptograma obtido: 18: 3_Texto3.txt
criptograma obtido: 19: 3_Texto4.txt
criptograma obtido: 20: 3_Texto5.txt
criptograma obtido: 21: 3_Texto6.txt
criptograma obtido: 22: 3_Texto7.txt
criptograma obtido: 23: 3_Texto8.txt
criptograma obtido: 24: 3_Texto9.txt
Quantidade de criptogramas obtidos pelo neuronio 71: 0
Quantidade de criptogramas obtidos pelo neuronio 72: 0
Quantidade de criptogramas obtidos pelo neuronio 73: 0
Quantidade de criptogramas obtidos pelo neuronio 74: 0
Quantidade de criptogramas obtidos pelo neuronio 75: 0
Quantidade de criptogramas obtidos pelo neuronio 76: 0
Quantidade de criptogramas obtidos pelo neuronio 77: 0
Quantidade de criptogramas obtidos pelo neuronio 78: 0
Quantidade de criptogramas obtidos pelo neuronio 79: 0

Quantidade de criptogramas obtidos pelo neuronio 710: 0
Quantidade de criptogramas obtidos pelo neuronio 81: 0
Quantidade de criptogramas obtidos pelo neuronio 82: 0
Quantidade de criptogramas obtidos pelo neuronio 83: 0
Quantidade de criptogramas obtidos pelo neuronio 84: 0
Quantidade de criptogramas obtidos pelo neuronio 85: 0
Quantidade de criptogramas obtidos pelo neuronio 86: 0
Quantidade de criptogramas obtidos pelo neuronio 87: 0
Quantidade de criptogramas obtidos pelo neuronio 88: 0
Quantidade de criptogramas obtidos pelo neuronio 89: 0
Quantidade de criptogramas obtidos pelo neuronio 810: 0
Quantidade de criptogramas obtidos pelo neuronio 91: 0
Quantidade de criptogramas obtidos pelo neuronio 92: 0
Quantidade de criptogramas obtidos pelo neuronio 93: 0
Quantidade de criptogramas obtidos pelo neuronio 94: 0
Quantidade de criptogramas obtidos pelo neuronio 95: 0
Quantidade de criptogramas obtidos pelo neuronio 96: 0
Quantidade de criptogramas obtidos pelo neuronio 97: 1
criptograma obtido: 1: 2_Texto1.txt
Quantidade de criptogramas obtidos pelo neuronio 98: 1
criptograma obtido: 1: 7_Texto1.txt
Quantidade de criptogramas obtidos pelo neuronio 99: 0
Quantidade de criptogramas obtidos pelo neuronio 910: 0
Quantidade de criptogramas obtidos pelo neuronio 101: 0
Quantidade de criptogramas obtidos pelo neuronio 102: 0
Quantidade de criptogramas obtidos pelo neuronio 103: 0
Quantidade de criptogramas obtidos pelo neuronio 104: 0
Quantidade de criptogramas obtidos pelo neuronio 105: 1
criptograma obtido: 1: 1_Texto1.txt
Quantidade de criptogramas obtidos pelo neuronio 106: 3
criptograma obtido: 1: 1_Texto10.txt
criptograma obtido: 2: 10_Texto1.txt
criptograma obtido: 3: 2_Texto10.txt
Quantidade de criptogramas obtidos pelo neuronio 107: 23
criptograma obtido: 1: 1_Texto11.txt
criptograma obtido: 2: 1_Texto12.txt
criptograma obtido: 3: 1_Texto13.txt
criptograma obtido: 4: 1_Texto14.txt
criptograma obtido: 5: 1_Texto18.txt
criptograma obtido: 6: 1_Texto19.txt
criptograma obtido: 7: 1_Texto2.txt
criptograma obtido: 8: 1_Texto20.txt
criptograma obtido: 9: 1_Texto21.txt
criptograma obtido: 10: 1_Texto22.txt
criptograma obtido: 11: 1_Texto23.txt
criptograma obtido: 12: 1_Texto24.txt
criptograma obtido: 13: 1_Texto5.txt
criptograma obtido: 14: 1_Texto6.txt
criptograma obtido: 15: 1_Texto7.txt

criptograma obtido: 16: 1_Texto8.txt
criptograma obtido: 17: 1_Texto9.txt
criptograma obtido: 18: 10_Texto10.txt
criptograma obtido: 19: 10_Texto11.txt
criptograma obtido: 20: 2_Texto11.txt
criptograma obtido: 21: 2_Texto12.txt
criptograma obtido: 22: 4_Texto1.txt
criptograma obtido: 23: 6_Texto1.txt
Quantidade de criptogramas obtidos pelo neuronio 108: 38
criptograma obtido: 1: 1_Texto15.txt
criptograma obtido: 2: 1_Texto16.txt
criptograma obtido: 3: 1_Texto17.txt
criptograma obtido: 4: 1_Texto3.txt
criptograma obtido: 5: 1_Texto4.txt
criptograma obtido: 6: 10_Texto12.txt
criptograma obtido: 7: 10_Texto13.txt
criptograma obtido: 8: 10_Texto14.txt
criptograma obtido: 9: 10_Texto15.txt
criptograma obtido: 10: 10_Texto16.txt
criptograma obtido: 11: 10_Texto17.txt
criptograma obtido: 12: 10_Texto18.txt
criptograma obtido: 13: 10_Texto19.txt
criptograma obtido: 14: 10_Texto2.txt
criptograma obtido: 15: 10_Texto20.txt
criptograma obtido: 16: 10_Texto21.txt
criptograma obtido: 17: 10_Texto22.txt
criptograma obtido: 18: 10_Texto23.txt
criptograma obtido: 19: 10_Texto24.txt
criptograma obtido: 20: 10_Texto3.txt
criptograma obtido: 21: 10_Texto5.txt
criptograma obtido: 22: 10_Texto6.txt
criptograma obtido: 23: 10_Texto7.txt
criptograma obtido: 24: 10_Texto8.txt
criptograma obtido: 25: 10_Texto9.txt
criptograma obtido: 26: 2_Texto13.txt
criptograma obtido: 27: 2_Texto14.txt
criptograma obtido: 28: 2_Texto15.txt
criptograma obtido: 29: 2_Texto16.txt
criptograma obtido: 30: 2_Texto17.txt
criptograma obtido: 31: 2_Texto18.txt
criptograma obtido: 32: 2_Texto20.txt
criptograma obtido: 33: 2_Texto22.txt
criptograma obtido: 34: 4_Texto10.txt
criptograma obtido: 35: 4_Texto11.txt
criptograma obtido: 36: 4_Texto12.txt
criptograma obtido: 37: 6_Texto10.txt
criptograma obtido: 38: 7_Texto10.txt
Quantidade de criptogramas obtidos pelo neuronio 109: 19
criptograma obtido: 1: 10_Texto4.txt

criptograma obtido: 2: 2_Texto19.txt
criptograma obtido: 3: 2_Texto2.txt
criptograma obtido: 4: 2_Texto21.txt
criptograma obtido: 5: 2_Texto23.txt
criptograma obtido: 6: 2_Texto24.txt
criptograma obtido: 7: 2_Texto3.txt
criptograma obtido: 8: 2_Texto5.txt
criptograma obtido: 9: 2_Texto6.txt
criptograma obtido: 10: 2_Texto7.txt
criptograma obtido: 11: 2_Texto8.txt
criptograma obtido: 12: 2_Texto9.txt
criptograma obtido: 13: 4_Texto13.txt
criptograma obtido: 14: 4_Texto14.txt
criptograma obtido: 15: 6_Texto11.txt
criptograma obtido: 16: 7_Texto11.txt
criptograma obtido: 17: 7_Texto12.txt
criptograma obtido: 18: 7_Texto13.txt
criptograma obtido: 19: 7_Texto14.txt
Quantidade de criptogramas obtidos pelo neuronio 1010: 82
criptograma obtido: 1: 2_Texto4.txt
criptograma obtido: 2: 4_Texto15.txt
criptograma obtido: 3: 4_Texto16.txt
criptograma obtido: 4: 4_Texto17.txt
criptograma obtido: 5: 4_Texto18.txt
criptograma obtido: 6: 4_Texto19.txt
criptograma obtido: 7: 4_Texto2.txt
criptograma obtido: 8: 4_Texto20.txt
criptograma obtido: 9: 4_Texto21.txt
criptograma obtido: 10: 4_Texto22.txt
criptograma obtido: 11: 4_Texto23.txt
criptograma obtido: 12: 4_Texto24.txt
criptograma obtido: 13: 4_Texto3.txt
criptograma obtido: 14: 4_Texto4.txt
criptograma obtido: 15: 4_Texto5.txt
criptograma obtido: 16: 4_Texto6.txt
criptograma obtido: 17: 4_Texto7.txt
criptograma obtido: 18: 4_Texto8.txt
criptograma obtido: 19: 4_Texto9.txt
criptograma obtido: 20: 6_Texto12.txt
criptograma obtido: 21: 6_Texto13.txt
criptograma obtido: 22: 6_Texto14.txt
criptograma obtido: 23: 6_Texto15.txt
criptograma obtido: 24: 6_Texto16.txt
criptograma obtido: 25: 6_Texto17.txt
criptograma obtido: 26: 6_Texto18.txt
criptograma obtido: 27: 6_Texto19.txt
criptograma obtido: 28: 6_Texto2.txt
criptograma obtido: 29: 6_Texto20.txt
criptograma obtido: 30: 6_Texto21.txt

criptograma obtido: 31: 6_Texto22.txt
criptograma obtido: 32: 6_Texto23.txt
criptograma obtido: 33: 6_Texto24.txt
criptograma obtido: 34: 6_Texto3.txt
criptograma obtido: 35: 6_Texto4.txt
criptograma obtido: 36: 6_Texto5.txt
criptograma obtido: 37: 6_Texto6.txt
criptograma obtido: 38: 6_Texto7.txt
criptograma obtido: 39: 6_Texto8.txt
criptograma obtido: 40: 6_Texto9.txt
criptograma obtido: 41: 7_Texto15.txt
criptograma obtido: 42: 7_Texto16.txt
criptograma obtido: 43: 7_Texto17.txt
criptograma obtido: 44: 7_Texto18.txt
criptograma obtido: 45: 7_Texto19.txt
criptograma obtido: 46: 7_Texto2.txt

9.20.2 FASE DE TESTE (CLASSIFICAÇÃO)

Quantidade de criptogramas obtidos pelo neuronio 11: 4
criptograma obtido: 1: 8_Texto25.txt
criptograma obtido: 2: 8_Texto26.txt
criptograma obtido: 3: 8_Texto27.txt
criptograma obtido: 4: 8_Texto29.txt
Quantidade de criptogramas obtidos pelo neuronio 12: 1
criptograma obtido: 1: 8_Texto28.txt
Quantidade de criptogramas obtidos pelo neuronio 13: 1
criptograma obtido: 1: 8_Texto30.txt
Quantidade de criptogramas obtidos pelo neuronio 14: 0
Quantidade de criptogramas obtidos pelo neuronio 15: 0
Quantidade de criptogramas obtidos pelo neuronio 16: 0
Quantidade de criptogramas obtidos pelo neuronio 17: 0
Quantidade de criptogramas obtidos pelo neuronio 18: 0
Quantidade de criptogramas obtidos pelo neuronio 19: 0
Quantidade de criptogramas obtidos pelo neuronio 110: 0
Quantidade de criptogramas obtidos pelo neuronio 21: 0
Quantidade de criptogramas obtidos pelo neuronio 22: 0
Quantidade de criptogramas obtidos pelo neuronio 23: 0
Quantidade de criptogramas obtidos pelo neuronio 24: 0
Quantidade de criptogramas obtidos pelo neuronio 25: 0
Quantidade de criptogramas obtidos pelo neuronio 26: 0
Quantidade de criptogramas obtidos pelo neuronio 27: 0
Quantidade de criptogramas obtidos pelo neuronio 28: 0
Quantidade de criptogramas obtidos pelo neuronio 29: 0
Quantidade de criptogramas obtidos pelo neuronio 210: 6
criptograma obtido: 1: 5_Texto25.txt
criptograma obtido: 2: 5_Texto26.txt

criptograma obtido: 3: 5_Texto27.txt
criptograma obtido: 4: 5_Texto28.txt
criptograma obtido: 5: 5_Texto29.txt
criptograma obtido: 6: 5_Texto30.txt
Quantidade de criptogramas obtidos pelo neuronio 31: 0
Quantidade de criptogramas obtidos pelo neuronio 32: 0
Quantidade de criptogramas obtidos pelo neuronio 33: 0
Quantidade de criptogramas obtidos pelo neuronio 34: 0
Quantidade de criptogramas obtidos pelo neuronio 35: 0
Quantidade de criptogramas obtidos pelo neuronio 36: 0
Quantidade de criptogramas obtidos pelo neuronio 37: 0
Quantidade de criptogramas obtidos pelo neuronio 38: 0
Quantidade de criptogramas obtidos pelo neuronio 39: 0
Quantidade de criptogramas obtidos pelo neuronio 310: 0
Quantidade de criptogramas obtidos pelo neuronio 41: 0
Quantidade de criptogramas obtidos pelo neuronio 42: 0
Quantidade de criptogramas obtidos pelo neuronio 43: 0
Quantidade de criptogramas obtidos pelo neuronio 44: 0
Quantidade de criptogramas obtidos pelo neuronio 45: 0
Quantidade de criptogramas obtidos pelo neuronio 46: 0
Quantidade de criptogramas obtidos pelo neuronio 47: 0
Quantidade de criptogramas obtidos pelo neuronio 48: 0
Quantidade de criptogramas obtidos pelo neuronio 49: 0
Quantidade de criptogramas obtidos pelo neuronio 410: 0
Quantidade de criptogramas obtidos pelo neuronio 51: 0
Quantidade de criptogramas obtidos pelo neuronio 52: 0
Quantidade de criptogramas obtidos pelo neuronio 53: 0
Quantidade de criptogramas obtidos pelo neuronio 54: 0
Quantidade de criptogramas obtidos pelo neuronio 55: 0
Quantidade de criptogramas obtidos pelo neuronio 56: 0
Quantidade de criptogramas obtidos pelo neuronio 57: 0
Quantidade de criptogramas obtidos pelo neuronio 58: 0
Quantidade de criptogramas obtidos pelo neuronio 59: 0
Quantidade de criptogramas obtidos pelo neuronio 510: 0
Quantidade de criptogramas obtidos pelo neuronio 61: 0
Quantidade de criptogramas obtidos pelo neuronio 62: 0
Quantidade de criptogramas obtidos pelo neuronio 63: 0
Quantidade de criptogramas obtidos pelo neuronio 64: 0
Quantidade de criptogramas obtidos pelo neuronio 65: 0
Quantidade de criptogramas obtidos pelo neuronio 66: 0
Quantidade de criptogramas obtidos pelo neuronio 67: 0
Quantidade de criptogramas obtidos pelo neuronio 68: 0
Quantidade de criptogramas obtidos pelo neuronio 69: 0
Quantidade de criptogramas obtidos pelo neuronio 610: 6
criptograma obtido: 1: 3_Texto25.txt
criptograma obtido: 2: 3_Texto26.txt
criptograma obtido: 3: 3_Texto27.txt
criptograma obtido: 4: 3_Texto28.txt
criptograma obtido: 5: 3_Texto29.txt

criptograma obtido: 6: 3_Texto30.txt
Quantidade de criptogramas obtidos pelo neuronio 71: 0
Quantidade de criptogramas obtidos pelo neuronio 72: 0
Quantidade de criptogramas obtidos pelo neuronio 73: 0
Quantidade de criptogramas obtidos pelo neuronio 74: 0
Quantidade de criptogramas obtidos pelo neuronio 75: 0
Quantidade de criptogramas obtidos pelo neuronio 76: 0
Quantidade de criptogramas obtidos pelo neuronio 77: 0
Quantidade de criptogramas obtidos pelo neuronio 78: 0
Quantidade de criptogramas obtidos pelo neuronio 79: 0
Quantidade de criptogramas obtidos pelo neuronio 710: 0
Quantidade de criptogramas obtidos pelo neuronio 81: 0
Quantidade de criptogramas obtidos pelo neuronio 82: 0
Quantidade de criptogramas obtidos pelo neuronio 83: 0
Quantidade de criptogramas obtidos pelo neuronio 84: 0
Quantidade de criptogramas obtidos pelo neuronio 85: 0
Quantidade de criptogramas obtidos pelo neuronio 86: 0
Quantidade de criptogramas obtidos pelo neuronio 87: 0
Quantidade de criptogramas obtidos pelo neuronio 88: 0
Quantidade de criptogramas obtidos pelo neuronio 89: 0
Quantidade de criptogramas obtidos pelo neuronio 810: 0
Quantidade de criptogramas obtidos pelo neuronio 91: 0
Quantidade de criptogramas obtidos pelo neuronio 92: 0
Quantidade de criptogramas obtidos pelo neuronio 93: 0
Quantidade de criptogramas obtidos pelo neuronio 94: 0
Quantidade de criptogramas obtidos pelo neuronio 95: 0
Quantidade de criptogramas obtidos pelo neuronio 96: 0
Quantidade de criptogramas obtidos pelo neuronio 97: 0
Quantidade de criptogramas obtidos pelo neuronio 98: 0
Quantidade de criptogramas obtidos pelo neuronio 99: 0
Quantidade de criptogramas obtidos pelo neuronio 910: 0
Quantidade de criptogramas obtidos pelo neuronio 101: 0
Quantidade de criptogramas obtidos pelo neuronio 102: 0
Quantidade de criptogramas obtidos pelo neuronio 103: 0
Quantidade de criptogramas obtidos pelo neuronio 104: 0
Quantidade de criptogramas obtidos pelo neuronio 105: 0
Quantidade de criptogramas obtidos pelo neuronio 106: 0
Quantidade de criptogramas obtidos pelo neuronio 107: 6
criptograma obtido: 1: 1_Texto25.txt
criptograma obtido: 2: 1_Texto26.txt
criptograma obtido: 3: 1_Texto27.txt
criptograma obtido: 4: 1_Texto28.txt
criptograma obtido: 5: 1_Texto29.txt
criptograma obtido: 6: 1_Texto30.txt
Quantidade de criptogramas obtidos pelo neuronio 108: 12
criptograma obtido: 1: 10_Texto25.txt
criptograma obtido: 2: 10_Texto26.txt
criptograma obtido: 3: 10_Texto27.txt
criptograma obtido: 4: 10_Texto28.txt

criptograma obtido: 5: 10_Texto29.txt
criptograma obtido: 6: 10_Texto30.txt
criptograma obtido: 7: 2_Texto25.txt
criptograma obtido: 8: 2_Texto26.txt
criptograma obtido: 9: 2_Texto27.txt
criptograma obtido: 10: 2_Texto28.txt
criptograma obtido: 11: 2_Texto29.txt
criptograma obtido: 12: 2_Texto30.txt
Quantidade de criptogramas obtidos pelo neuronio 109: 0
Quantidade de criptogramas obtidos pelo neuronio 1010: 24
criptograma obtido: 1: 4_Texto25.txt
criptograma obtido: 2: 4_Texto26.txt
criptograma obtido: 3: 4_Texto27.txt
criptograma obtido: 4: 4_Texto28.txt
criptograma obtido: 5: 4_Texto29.txt
criptograma obtido: 6: 4_Texto30.txt
criptograma obtido: 7: 6_Texto25.txt
criptograma obtido: 8: 6_Texto26.txt
criptograma obtido: 9: 6_Texto27.txt
criptograma obtido: 10: 6_Texto28.txt
criptograma obtido: 11: 6_Texto29.txt
criptograma obtido: 12: 6_Texto30.txt
criptograma obtido: 13: 7_Texto25.txt
criptograma obtido: 14: 7_Texto26.txt
criptograma obtido: 15: 7_Texto27.txt
criptograma obtido: 16: 7_Texto28.txt
criptograma obtido: 17: 7_Texto29.txt
criptograma obtido: 18: 7_Texto30.txt
criptograma obtido: 19: 9_Texto25.txt
criptograma obtido: 20: 9_Texto26.txt
criptograma obtido: 21: 9_Texto27.txt
criptograma obtido: 22: 9_Texto28.txt
criptograma obtido: 23: 9_Texto29.txt
criptograma obtido: 24: 9_Texto30.txt

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo