



## **Relatório Final**

### **PIBIC & PIBIC-AF, PIBITI E VOLUNTÁRIOS**

<Observação: Favor não alterar o layout desta página de rosto. Apenas preencha os dados nos campos solicitados. A partir da segunda página estão os demais itens do modelo a serem preenchidos.>

EDITAL / PROGRAMA
<b>EDITAL PROPCI/UFBA_01_PIBIC 2015-2016</b>

ESTUDANTE IC (Digitar nome completo, sem abreviações).
<b>Nilton Vasques Carvalho Junior</b>

Título do Plano de Trabalho do Estudante (Digitar o título completo, sem abreviações, exatamente igual ao título do plano de trabalho aprovado).
<b>Extração de padrões para gerenciamento de imprecisão e incerteza na organização flexível de documentos de textos</b>

ORIENTADOR (Digitar nome completo, sem abreviações).
<b>Tatiane Nogueira Rios</b>

Título do Projeto do Orientador (Digitar o título completo, sem abreviações, exatamente igual ao título do projeto do orientador).
<b>Gerenciamento de imprecisão e incerteza para organização flexível de documentos de texto utilizando a teoria de conjuntos fuzzy</b>

Salvador  
Agosto de 2016



## RESUMO

Este relatório visa apresentar uma investigação dos impactos dos algoritmos de agrupamento na composição da organização flexível de documentos textuais. Discutindo aspectos relevantes da etapa de pré-processamento dos documentos a serem organizados; da realização de agrupamento para extração de padrões nestes documentos e dos critérios de validação da organização obtida. A pesquisa realizada para desenvolvimento deste trabalho tem como principal contribuição uma abordagem híbrida para a organização flexível de documentos, mesclando os benefícios concedidos pela adequada interpretação de partições fuzzy e possibilísticas existentes no método de agrupamento Possibilistic C-Means e Possibilistic Fuzzy C-Means (PFCM). Como fruto das análises aqui realizadas, foram propostos os métodos de extração de descritores Possibilistic Description Comes Last (PDCL) e Mixed - Possibilistic Fuzzy Description Comes Last (PFDCL). Ambos mostraram-se relevantes através de evidências experimentais e análises subjetivas à adequação dos métodos, para a organização flexível de documentos, contribuindo com descobertas originais para o estado da arte. Os resultados da pesquisa ainda estimulam novas implementações cuja execução pode transcorrer em trabalhos futuros.

## 1. INTRODUÇÃO

O avanço da computação pessoal, em particular a computação móvel, tem proporcionado um gigantesco aumento da quantidade de dados armazenados pela humanidade ao longo dos anos. A critério de exemplo, a popular plataforma de rede social Facebook <sup>1</sup>, produz diariamente mais de 25 terabytes de informação [1]. A tendência com o avanço das tecnologias, é que tudo seja integrado a internet, de tal modo que os pesquisadores já chegam a dizer que os dados são o novo recurso natural do planeta [2]. Ainda segundo os autores, entre as maiores fontes de geração de dados estão os sistemas governamentais, plataformas de mídias sociais, assim como arquivos armazenados pelas corporações, como por exemplo, formulários médicos, opiniões de consumidores, relatórios e etc.. Entretanto, é importante ressaltar que todos esses dados excedem os limites humanos para o uso e compreensão destes [3].

Diante desse cenário, instituições públicas e privadas estão sobrecarregadas com a tarefa de processar essa imensa quantidade de informação em bases de dados com documentos não estruturados e em diversos formatos [4]. Estes documentos usualmente são de diversos tipos, como por exemplo, textos, áudios, imagens, vídeos, documentos HTML, podendo estar, inclusive, em diferentes idiomas.

Nesse contexto, diversas pesquisas tem objetivado a proposição ou refinamento de técnicas para automatização do processo de análise e aquisição de conhecimento útil desse montante de informações armazenadas. Porém, devido a multi disciplinaridade inerente desse campo de estudo, o mesmo tem sido estudado pelas comunidades de mineração de dados, aprendizado de máquina e recuperação de informação.

A Mineração de Dados (MD) é um campo de estudo que vem obtendo rápidos avanços nos últimos anos, e, isto se deve aos avanços das tecnologias de hardware e software, o qual possibilitou o massivo armazenamento de diferentes tipos de dados, inclusive os dados textuais [5]. Portanto, como resultado desse aumento na quantidade de documentos disponíveis na forma textual, existe uma demanda crescente no desenvolvimento e aprimoramento de métodos e algoritmos que possam efetivamente processar e extrair padrões dos dados de maneira dinâmica e escalável.

Por outro lado, enquanto os dados estruturados já possuem mecanismos bem eficientes de armazenamento e recuperação, os dados textuais são geralmente gerenciados através de mecanismos de buscas para suprir essa falta de estruturação. Esses mecanismos de busca possibilitam aos usuários uma conveniente maneira para recuperar informações em coleções textuais através de consultas com palavras chaves. Desse modo, compete



ao campo de estudo da Recuperação de Informação (RI) a tarefa de explorar, investigar e propor métodos para otimização da eficiência e efetividade de sistemas de buscas [6].

Mas segundo Baeza-Yates e Ribeiro-Neto (2011), as pesquisas de recuperação de informação tem focado tradicionalmente, em formas de facilitar o acesso à informação, do que realizar a descoberta de novos padrões em documentos textuais, o qual se destaca como sendo o objetivo principal da mineração de textos. A mineração de textos, por sua vez, é uma especialização da mineração de dados, que busca incorporar atividades de estruturação dos documentos em formatos apropriados, facilitando a aplicação dos tradicionais métodos de extração de padrões da MD, minimizando as perdas durante a conversão do formato original não estruturado [7].

Contudo, uma série de características diferenciam os documentos textuais de outras formas de dados. O que por sua vez, afeta o desempenho das clássicas técnicas da MD. Dentre essas características peculiares, destacam-se como mais importantes os fatos de que os dados são esparsos e possuem alta dimensionalidade. Por exemplo, uma coleção de documentos (corpus) pode conter 100.000 palavras (termos), enquanto um único documento desse corpus pode conter somente algumas centenas de palavras [5]. Essa discrepância, tem implicações diretas em várias técnicas de identificação de padrões, e especialmente no agrupamento textual, que deriva de clássicas técnicas de agrupamento da mineração de dados, aplicados à conjuntos de baixa dimensionalidade.

Para cumprir a tarefa de extrair informações relevantes de documentos textuais e identificar as estruturas inerentes aos mesmos. A mineração de textos emprega uma variedade de técnicas, as quais se destacam aquelas usualmente desenvolvidas para efetuar tarefas de coleta, pré-processamento, agrupamento textual e seleção de termos descritores para o agrupamento.

O agrupamento pode, de maneira geral, ser definido como a tarefa de agrupar uma coleção de objetos, de acordo algum critério de similaridade. É possível distinguir os tipos de agrupamento em função da lógica empregada por eles. Com isso, tem-se os algoritmos que derivam da lógica clássica ou da lógica fuzzy. Na lógica clássica, após a conclusão do agrupamento, cada elemento só pertence à apenas um grupo, enquanto que na lógica fuzzy, a pertinência do elemento será distribuída entre os grupos.

Ao se analisar a diversidade de conteúdo em dados textuais, é trivial notar que frequentemente um texto aborda um ou mais temas. O que implica que o agrupamento clássico, ao atribuir um objeto a apenas um grupo, não irá representar bem a imprecisão e incerteza natural dos documentos.

Deste modo, os métodos de agrupamento derivados da lógica fuzzy se mostram mais capacitados para lidar com essa imprecisão e incerteza da realidade multi temática dos documentos textuais. Assim sendo, uma organização flexível de documentos pode ser

definida como o processo que compreende a estruturação dos dados, a adição de flexibilidade proporcionada pelo agrupamento fuzzy, a extração de descritores dos grupos de maneira flexível e a recuperação de informação através de um Sistema de Recuperação de Informação (SRI).

Ao se observar o processo de organização flexível de documentos, percebe-se que o mesmo abrange várias etapas, cada uma delas com suas particularidades. No entanto, apesar da importância desempenhada por cada etapa do processo, o agrupamento em si pode ser visto como uma das peças chaves, pois ele é diretamente responsável por organizar os documentos de acordo com as suas similaridades. Adicionalmente, é preciso desconsiderar ou reduzir a influência de documentos ruidosos, que destoam do restante da coleção nos grupos finais.

O algoritmo Fuzzy C-Means (FCM) [8], que deriva do clássico K-Means [9], e o Possibilistic C-Means (PCM) [10], são exemplos de métodos de agrupamento capazes de organizar de maneira automatizada uma coleção de documentos em um conjunto de grupos. Ambos distribuem os documentos de uma coleção textual em um conjunto de grupos, de modo que cada documento possa pertencer a diferentes grupos com diferentes graus de pertinência, considerando assim a flexibilidade necessária para tratar a imprecisão e incerteza do processo.

No entanto, o FCM apresenta alguns resultados indesejados, diante da presença de



dados ruidosos na coleção. Em se tratando de coleções textuais, um dado ruidoso pode ser considerado como um documento que possua uma temática bastante diferente dos demais documentos da coleção. Com o objetivo de atribuir valores de pertinências mais realísticos aos elementos a serem agrupados e penalizar com baixas pertinências os elementos ruidosos, o algoritmo PCM foi proposto. Porém, o PCM é muito sensível à inicialização, o que pode resultar em grupos coincidentes, onde não há uma separação muito bem definida dos elementos.

Visando contemplar os benefícios de ambos os métodos, foi proposto o método de agrupamento Possibilistic Fuzzy C-Means (PFCM) [11], como uma versão híbrida dos algoritmos FCM e o PCM, objetivando adicionar robustez à tarefa de agrupamento.

Após o agrupamento dos documentos, é necessário realizar a extração dos termos que melhor descrevem os grupos. Para realização dessa tarefa, tem-se alguns métodos na literatura do tipo DCF (Description Comes First), que realizam a extração de maneira embutida no processo de agrupamento. Porém, essa abordagem torna o processo de extração de descritores dependente do algoritmo de agrupamento. Com o propósito de contornar esse cenário, foi proposto em Nogueira (2013) o método Soft Organization Fuzzy Description Comes Last (SoftO-FDCL) (Nogueira, 2013), o qual extrai os termos descritores após a etapa de agrupamento de maneira independente do algoritmo de agrupamento utilizado. Permitindo avaliar diretamente os impactos dos métodos de agrupamento, na extração de descritores e por consequência na qualidade da organização flexível de documentos.

Entretanto, o método SoftO-FDCL foi pensado inicialmente para interpretar as pertinências produzidas na partição do FCM, que difere da partição resultante produzida pelo PCM. A principal contribuição do PCM foi uma alteração no modo de atribuição da pertinência de um elemento a um grupo, o que impacta diretamente na partição dos grupos resultantes.

Diante deste contexto, e tendo em vista o crescente aumento de informações produzidas além da capacidade humana de analisar. Com a demanda crescente no desenvolvimento e aprimoramento das técnicas de extração e identificação de conhecimento útil em dados textuais, bem como a necessidade de se organizar esses dados de maneira flexível, tratando a imprecisão e incerteza natural desses dados e considerando as particularidades existentes nos métodos de agrupamento, foi formulada a seguinte hipótese para o desenvolvimento desse trabalho:

*A utilização de uma estratégia híbrida de agrupamento e extração de descritores, entre os graus de pertinência e tipicidade providos pelo método de agrupamento PFCM, permitem o aumento da robustez e resiliência contra ruídos na organização flexível de documentos, aumentando assim a relevância dos grupos obtidos.*

Para demonstrar a validade da hipótese formulada, com base na exploração de estratégias existentes na literatura para o aprimoramento do processo de organização flexível de documentos, definiu-se o seguinte objetivo:

*Conduzir uma investigação em torno dos métodos de agrupamento FCM, PCM e PFCM, para compreender e interpretar corretamente as peculiaridades de se extrair descritores a partir de um agrupamento híbrido.*

Considerando-se os resultados dos experimentos realizados, foi descoberto que as alterações existentes no PCM, impactam diretamente na qualidade dos descritores extraídos pelo método SoftO-FDCL. Essa descoberta motivou a proposição de dois novos métodos de extração de descritores: Possibilistic Descriptor Comes Last (PDCL) e Mixed - Possibilistic Fuzzy Descriptor Comes Last (Mixed-PFDCL). Os quais apresentaram resultados que contribuem de maneira significativa para o estado da arte da extração de descritores dos grupos fuzzy e para o aprimoramento da organização flexível de documentos.



## 2. MATERIAIS E MÉTODOS

Entre os principais fundamentos necessários para compreensão da abordagem apresentada neste documento, estão a atividade de pré-processamento dos dados, cuja finalidade é filtrar e estruturar os dados para serem processados nas etapas seguintes; os principais algoritmos de agrupamento fuzzy, presentes na literatura, com as suas definições matemáticas e pseudo códigos; e, por fim, a tarefa de rotular os grupos encontrados na etapa de agrupamento com os termos que melhor os representem, permitindo assim a realização de consultas em Sistemas de Recuperação da Informação (SRI).

### 2.1 PRÉ PROCESSAMENTO

Pré-processamento dos dados é o processo de limpeza e preparação dos dados para extração de padrões. Para este trabalho, especificamente, considera-se dado como sendo um documento textual e a tarefa de extração de padrões a ser considerada é o agrupamento.

Essa etapa é importante porque algumas palavras em um documento podem causar pouco ou nenhum impacto no significado geral do documento [12]. Soma-se a isso o enorme custo computacional do processo de mineração de textos, devido à grande quantidade de atributos presente em dados textuais, visto que quanto maior for a coleção de textos, maior será a quantidade de palavras distintas. Tal dimensionalidade eleva bastante o custo computacional de qualquer tarefa de extração de padrões. Por isso, vários pesquisadores propuseram métodos para tentar simplificar, sintetizar e eliminar redundâncias desnecessárias nas coleções de textos.

O processo de pré-processamento de dados textuais, inicia com um documento parcialmente estruturado e avança incrementando a estrutura através do refinamento das características do documento e adicionando novas [13]. No contexto da mineração de textos, as características dos documentos são as suas palavras [12]. Ao final do processo, as palavras mais relevantes são utilizadas, e as demais são descartadas.

### 2.2 AGRUPAMENTO FUZZY

O agrupamento é um processo não supervisionado cujo o objetivo é organizar os objetos similares no mesmo grupo e os objetos com grau de dissimilaridade elevado em grupos distintos [7]. Este processo é de grande utilidade para diversos campos de estudo da inteligência computacional, como a mineração de dados, recuperação de informação, segmentação de imagens e classificação de padrões. Neste trabalho, os objetos a serem agrupados são os documentos textuais.

#### 2.2.1 Fuzzy C Means (FCM)

Bezdek et al. (1984) descreve um método de agrupamento fuzzy que produz como saída partições fuzzy e protótipos dos grupos. Esse algoritmo desempenha um papel importante no contexto do agrupamento fuzzy, devido ao seu pioneirismo no campo de estudo, possuindo diversas extensões, sendo considerado um dos mais amplamente utilizados métodos de agrupamento fuzzy da literatura [11]. A maioria dos métodos de agrupamento fuzzy são derivações do FCM [10].

O algoritmo fuzzy c means é baseado no algoritmo crisp K-means, com uma adaptação para a lógica fuzzy. Ou seja, ao contrário do algoritmo K-means, o FCM passa a permitir que um documento pertença a mais de um grupo. Nos experimentos, foi utilizada a versão iterativa do FCM, onde a cada iteração, a pseudo partição (matriz com as informações de pertinência dos documentos nos grupos) fuzzy é atualizada.

O FCM busca atualizar a pseudo partição fuzzy e minimizar a função objetivo apresentado na Figura 1, a cada iteração. O critério de parada é o parâmetro epsilon, que limita o grau de refinamento na partição.



$$J(P) = \sum_{i=1} \sum_{j=1} [u(di, gl)]^n ||di - vl||$$

Figura 1

Por conta da maneira como a função objetivo é definida, o FCM tem uma restrição que obriga que a soma das pertinências de um documento em todos os grupos seja sempre 1 (um). Isso leva ao problema dos elementos equidistantes, onde um documento que tenha a mesma similaridade entre os grupos, receberá a pertinência média, mesmo que a similaridade deste documento entre os grupos não seja relevante. A partição inicial do FCM é gerada distribuindo aleatoriamente a pertinência dos documentos nos grupos. Vale ressaltar que todos os demais algoritmos que foram utilizados nessa pesquisa, possuem essa característica, por serem extensões do FCM.

### 2.2.2 Possibilistic C Means (PCM)

A restrição probabilística do FCM, que obriga a soma das pertinências de um elemento ser igual a um, nem sempre resulta em pertinências que representam bem a realidade dos dados. Esse problema se agrava ainda mais, em bases com muitos dados ruidosos (outliers). Portanto, com o objetivo de contornar esses problemas do FCM, foi proposto em Krishnapuram e Keller (1993) o algoritmo Possibilistic C Means (PCM).

Ao contrário do FCM, o PCM não atribui pertinências dos documentos aos grupos, mas sim tipicidades, as quais podem ser interpretadas como graus de possibilidade de um elemento pertencer a um determinado grupo. Como consequência, a partição resultante é possibilística.

A tipicidade do documento em relação ao grupo, tem como objetivo ponderar os elementos que tenham mais relevância em relação ao grupo, e penalizar os documentos menos relevantes. De modo a remover completamente a restrição imposta pelo FCM. A função objetivo do FCM é então modificada pela equação (Figura 2).

$$K(P) = \sum_{i=1} \sum_{j=1} [\varphi(di, gl)]^n ||di - vl|| + \sum_{l=1} \gamma_l \sum_{i=1} (1 - \varphi(di, gl))^m$$

Figura 2

Apesar das melhorias propostas no PCM, existe também uma peculiaridade neste algoritmo, conhecida na literatura como problema dos clusters coincidentes [6]. Este problema ocorre, quando os grupos fuzzy são dispostos muito juntos, levando assim a uma alta interseção entre os conjuntos.

### 2.2.3 Possibilistic Fuzzy C Means (PFCM)

Com o propósito de aproveitar os benefícios de ambas as abordagens, Pal et al. (2005) propôs o algoritmo PFCM, que utiliza as pertinências do FCM e as tipicidades do PCM. Cabe ao usuário definir a proporção de cada uma das contribuições com parâmetros que ponderam o peso de ambos. Para tanto, é realizada uma mistura entre as funções objetivo apresentadas nas Figuras 1 e 2, resultando na minimização da função objetivo apresentada na Figura 3.

$$L(P) = \sum_{i=1} \sum_{j=1} [a[\phi(di, gl)]^n + b[\varphi(di, gl)]^m] ||di - vl|| + \sum_{l=1} \gamma_l \sum_{i=1} (1 - \varphi(di, gl))^m$$

Figura 3

## 2.3 EXTRAÇÃO DE DESCRITORES

A tarefa de atribuir significados à grupos é um dos problemas chave do agrupamento de



textos, pois ao final do processo de agrupamento, os grupos precisam apresentar alguma relevância para o usuário [14]. Portanto, é imprescindível que sejam extraídos descritores significativos para representar os documentos que compõem os grupos.

A etapa de extração de descritores etapa pode ser realizada manualmente, com o usuário guiando o processo, ou de forma automatizada, que por sua vez é mais interessante para a proposta de organização flexível de documentos, visto que, para grandes bases de dados textuais, a tarefa de extrair descritores para todos os grupos encontrados durante o agrupamento, pode ser bastante exaustiva para o usuário.

Dentre os métodos automatizados, dois tipos de abordagens são encontrados na literatura, uma baseada em conhecimento interno e a outra baseada em conhecimento externo. A primeira se utiliza somente de informações que podem ser obtidas na coleção de documentos, como por exemplo a frequência do termo, localização do termo na estrutura do documento. Enquanto a abordagem de conhecimento externo, leva em consideração fontes de informação externas, como por exemplo a consulta a extensa base de termos na língua inglesa WordNet, para auxiliar a escolha dos termos mais representativos.

Nogueira (2013) destaca que grande parte dos métodos de extração de descritores encontrados na literatura são embutidos na fase de agrupamento. O que justifica a avaliação dos mesmos em função do desempenho do agrupamento. No entanto, essa junção da extração de descritores na fase de agrupamento, dificulta a combinação de diferentes técnicas de agrupamento e consequentemente a escolha de bons descritores. Logo, os métodos onde a extração é realizada após a fase de agrupamento, de maneira independente, permitem uma melhor adaptação da proposta de organização flexível de documentos para diferentes contextos.

Nesse contexto, percebe-se à existência de algumas estratégias para extração de descritores, utilizando ou não conhecimento externo, e embutida ou independente do processo de agrupamento. A partir da avaliação dessas abordagens, e de acordo com o objetivo definido, considerou-se que a abordagem independente do algoritmo de agrupamento é mais pertinente ao presente estudo, pois ela viabiliza a condução de experimentos com vários métodos de agrupamento. Nesse sentido, foi escolhido o método Soft Organization - Fuzzy Description Comes last (SoftO-FDCL) proposto por Nogueira (2013), devido o mesmo possuir essas características necessárias para a investigação dos impactos do agrupamento na qualidade dos descritores e da organização flexível.

## 2.4 COLEÇÕES TEXTUAIS

Na mineração de textos e consequentemente nos trabalhos relacionados à organização flexível de documentos, é comum se realizar a avaliação dos métodos propostos, conduzindo se experimentos sobre coleções textuais existentes na literatura com essa finalidade. Para isso, as coleções precisam se estarem estruturadas. Assim sendo, nesta pesquisa foi adotada a estrutura de representação de documentos textuais tf-idf, como forma de estruturar os dados presentes nas coleções, de modo a capturar a importância relativa dos termos nos documentos e nas coleções, montando assim ao final do pré-processamento uma matriz documentos x termos.

Outro aspecto não menos importante, são as características particulares das coleções textuais. Pois ressalta-se que para uma mais apurada análise dos resultados, é pertinente considerar as particularidades de cada coleção, com a finalidade de encontrar possíveis justificativas para os resultados apresentados, realizando-se indagações comparativas às peculiaridades sabidamente conhecidas dos métodos analisados. O conjunto de características particulares de cada coleção, dar-se à como apresentado na Tabela 1. Uma análise objetiva das características presentes nas seis coleções utilizadas nos experimentos pode ser observada na Tabela 2, onde é possível notar de maneira bem objetiva ao se observar a coluna com o percentual de zeros da tabela, que todas as coleções apresentam uma quantidade de zeros em mais de 90% das frequências dos termos presentes na tf – idf, o que deixa explícito o peculiar problema dos dados esparsos já caracterizado ao longo do texto, como algo inerente aos dados textuais e que



afeta negativamente grande parte dos resultados do processo de mineração de textos.

<b>documentos</b>	número de documentos presentes na coleção
<b>termos</b>	número de termos existentes na coleção após o pré-processamento
<b>% zeros</b>	número relativo de zeros na <i>tf-idf</i> , quantificando o quanto a matriz é esparsa
<b>classes</b>	número de classes presentes na coleção
<b>n-gramas</b>	quantidade de termos considerados sequencialmente na coleção

Tabela 1: Descrição das características objetivas presentes em coleções textuais elenca-das para este trabalho

Coleção	docs	termos	classes	% zeros	n-gramas
Opinosis	51	842	3	95,73%	1-grama
20newsgroups	2000	11028	4	99,11%	1-grama
Hitech	600	6925	6	97,93%	1-grama
NSF	1600	2806	16	99,76%	1-grama
WAP	1560	8070	20	98,51%	1-grama
Reuters-21578	1052	3925	43	98,55%	1-grama

Tabela 2: Características das coleções textuais utilizadas nesta pesquisa

### 3. RESULTADOS

#### 3.1 REFINAMENTO COM O ALGORITMO PFCM

Conforme ficou evidenciado, a tarefa de organizar de maneira flexível um conjunto de documentos textuais, possui diversos desafios. Em particular, ao se agrupar um conjunto de documentos é esperado que os grupos resultantes possuam significado relevante, ou seja o algoritmo de agrupamento precisa detectar a estrutura natural dos documentos [15]. Alguns desses desafios estão na dificuldade em escalar os métodos usuais para coleções textuais de grande dimensionalidade; na obtenção de mecanismos efetivos para se avaliar a qualidade dos grupos produzidos; nas técnicas para se medir a interpretabilidade dos resultados; na capacidade para estimar os parâmetros dos algoritmos; na possibilidade de funcionar de maneira incremental, reduzindo o custo computacional durante a atualização dos grupos com novos dados; e, também na capacidade de continuar a produzir bons resultados em cenários compostos de documentos ruidosos.

Diante dos desafios propostos, e com a evidência de que é possível aprimorar os resultados ao se utilizar novas estratégias de agrupamento, a investigação apresentada nesta seção tem como objetivo analisar de qual forma a organização de documentos pode ser otimizada, ao aplicar na etapa de agrupamento uma estratégia que misture as

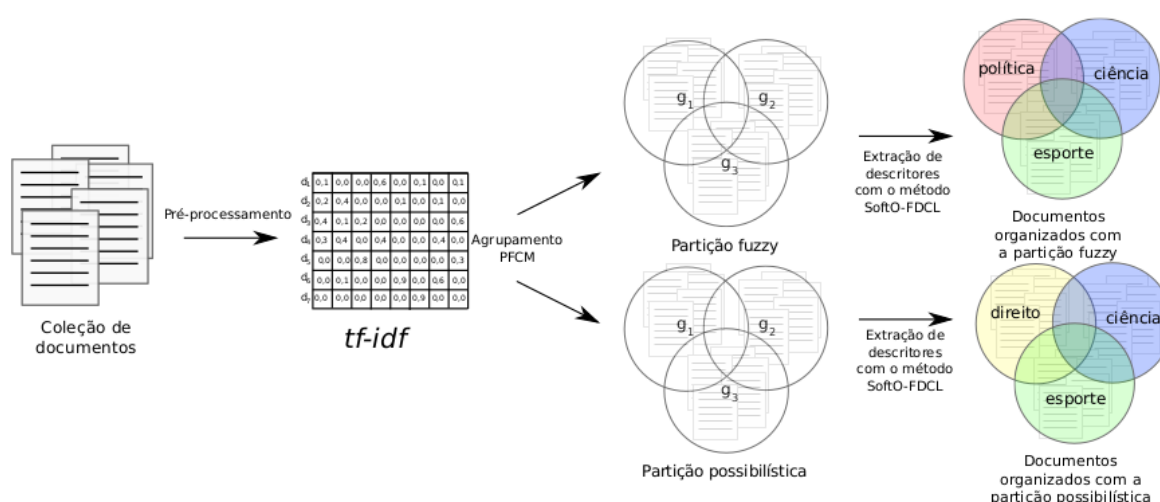


partições possibilística e fuzzy, por meio do algoritmo PFCM.

Vale ressaltar, que a escolha desse algoritmo foi feita devido o seu potencial para absorver as qualidades presentes no Fuzzy C-Means (FCM) contrabalanceando as suas deficiências ao agregar também o Possibilistic C-Means (PCM) e sua partição possibilística. Além disso, existem diversas pesquisas na literatura abordando o desempenho do PFCM, como por exemplo em [11].

Sendo assim, foram conduzidos experimentos adaptando a estratégia de organização flexível de documentos definida em Nogueira (2013), utilizando na etapa de agrupamento o método PFCM. Uma vez que esse método produz duas partições, uma possibilística e uma fuzzy, foi aplicado o método de extração de descritores SoftO-FDCL na partição fuzzy e também na partição possibilística, produzindo assim dois grupos de descritores.

Com essa adaptação espera-se uma melhor organização dos documentos, de forma que melhores descritores sejam escolhidos para caracterizar grupos. Tal processo de organização é ilustrado na Figura 4.



*Figura 4: Estratégia de organização flexível de documentos adotada ao se misturar abordagens fuzzy e possibilísticas no agrupamento.*

Para se calcular a quantidade ótima de grupos, para cada coleção foi utilizado o método da silhueta fuzzy, método bastante utilizado com o propósito de avaliar o agrupamento de documentos. Assim sendo, o número ideal de grupos é determinado após a execução da silhueta fuzzy variando o número de grupos entre 2 e o número de classes de cada coleção. Ressalta-se que em coleções que os documentos não possuem rótulos, ou seja o número de classes é desconhecido, ainda é possível usar o método da silhueta fuzzy para definir o número ótimo de grupos. No entanto, a quantidade máxima de grupos deve ser definida de modo empírico ou com base em alguma informação prévia a respeito dos dados.



Para permitir uma análise comparativa dos resultados, o experimento foi realizado também com os algoritmos FCM e PCM. Como resultado do agrupamento das coleções, está disposto na Tabela 3 a comparação do número de grupos obtidos por cada algoritmo de agrupamento. Nessa comparação nota-se que os algoritmos FCM e PFCM foram os que alcançaram uma quantidade de partições mais próxima da quantidade de classes existentes em cada coleção. Enquanto o PCM manteve uma tendência a produzir uma quantidade menor de grupos em relação aos demais.

Coleção	# classes	FCM	PCM	PFCM
Opinosis	3	3	3	3
20Newsgroup	4	2	2	2
Hitech	6	6	5	5
NSF	16	11	2	16
WAP	20	14	5	16
Reuters-21578	43	22	11	36

*Tabela 3: Quantidade ótima de grupos determinada através do método da silhueta fuzzy para cada algoritmo de agrupamento*

Após agrupar os dados utilizando os métodos FCM, PCM e PFCM, foi aplicado o método de extração de descritores SoftO-FDCL. Para avaliar os descritores produzidos, foi verificado o potencial preditivo dos mesmos, possibilitando assim quantificar a qualidade dos termos selecionados para nomear os grupos.

Visando avaliar a qualidade dos descritores e permitir uma comparação direta dos impactos dessa abordagem com os resultados publicados em Nogueira (2013). O resultado dos agrupamentos foi submetido aos algoritmos de classificação SVM, Naive Bayes, Multinomial Naive Bayes, KNN e C4.5, que são bem comuns na avaliação de métodos de aprendizado de máquina.

Nesse contexto, foi utilizada a implementação dos algoritmos de classificação anteriormente citados presentes na ferramenta WEKA. Os algoritmos Naive Bayes (NB), Multinomial Naive Bayes (NB-Multinomial) e o J48 (que é a implementação do C4.5 existente no WEKA), foram executados com os parâmetros padrão da ferramenta. Por outro lado, o SVM foi ajustado para usar o Normalized Polynomial Kernel com o parâmetro de complexidade sendo  $c = 2.0$ . O algoritmo IBk (implementação do KNN presente no WEKA) foi executado 7 vezes, variando o parâmetro de vizinhos de 1 até 7, sendo escolhido o melhor resultado. Ressalta-se que foi adotada a técnica 10-fold cross validation no experimento para melhor capturar a capacidade de generalização do modelo.



O resumo dos resultados do desempenho dos descritores extraídos após o agrupamento com os algoritmos FCM, PCM e PFCM é apresentado na Tabela 4. Na tabela, está marcado os métodos de agrupamento que obtiveram a maior taxa de classificação dentre os demais. Esses resultados obtidos reforçam a flexibilidade e adaptação do método SoftO-FDCL [7] a novos algoritmos de agrupamento, demonstrando-se promissor na tarefa de extrair termos relevantes dos grupos produzidos na etapa de agrupamento.

Coleção	docs	termos	classes	% zeros	FCM	PCM	PFCM
Opinosis	51	842	3	95,73%		✓	
20newsgroups	2000	11028	4	99,11%			✓
Hitech	600	6925	6	97,93%	✓		
NSF	1600	2806	16	99,76%	✓		
WAP	1560	8070	20	98,51%			✓
Reuters-21578	1052	3925	43	98,55%	✓		

Tabela 4: Sumário dos resultados da classificação dos descritores

Coleção	docs	termos	classes	% zeros	FCM	PCM	PFCM
Opinosis	51	842	3	95,73%		✓	
20newsgroups	2000	11028	4	99,11%			✓
Hitech	600	6925	6	97,93%	✓		
NSF	1600	2806	16	99,76%	✓		
WAP	1560	8070	20	98,51%			✓
Reuters-21578	1052	3925	43	98,55%	✓		

Tabela 5: Sumário dos resultados da classificação dos descritores

Adicionalmente, ressalta-se a importância também de avaliar de maneira subjetiva os descritores selecionados dos grupos, permitindo compreender se os termos obtidos fazem sentido para a organização de documentos em grupos.



Método	$crisp_1$	$crisp_2$	$crisp_3$
FCM	easy, clear, drive, display, control, car, version, nice, work, perfect	fact, import, isn't, model, problem, unit, design, don't, doesn't, found	breakfast, nearby, concierge, eat, bottle, coffee, floor, food, inn, friendly
PCM	easy, read, problem, version, don't, small, nice, car, work, found	fact, back, turn, expect, size, close, quality, review, min, feature	feel, amazing, isn't, extreme, drive, include, point, reason, give, run
PFCM $\mu$	easy, drive, control, don't, version, nice, car, work, perfect, lot	fact, isn't, read, complete, device, display, size, doesn't, found	breakfast, nearby, pleasant, concierge, eat, coffee, floor, clean, friendly, food
PFCM $\lambda$	club, immaculate, send, towel, basic, exception, spotl, pillow, typical, fridge	pub, housekeep, holiday, tourist, tea, smoke, pm, renovate, facilitate, london	usual, central, forum, bottle, modern, adult, supply, food, reserve, dinner

Table 6: Descritores extraídos com os métodos de agrupamento FCM, PCM e PFCM da coleção Opínosis, onde  $\mu$  e  $\lambda$  se referem as partições fuzzy e possibilística respectivamente, da qual os descritores foram extraídos.

Sendo assim, para uma análise subjetiva dos resultados, os descritores da coleção Opínosis, foram selecionados por possuírem poucos grupos facilitando a análise e a visualização. A coleção Opínosis contém opiniões dos usuários a respeito de serviços de hospedagem, dispositivos eletrônicos e carros e espera-se que os descritores de grupos se aproximem semanticamente de tais categorias. Na Tabela 4 tem-se a seleção de descritores escolhidos para cada grupo, extraídos pelos algoritmos FCM, PCM e PFCM. Ao analisar os descritores selecionados é possível notar uma tendência geral, do grupo  $crisp_1$  conter descritores relacionadas a carros, o  $crisp_2$  conter descritores sobre dispositivos eletrônicos e o grupo  $crisp_3$  descritores sobre hospedagem e alimentação. Contudo, nota-se que os descritores do PCM e do PFCM $\lambda$  (descritores da partição possibilística do PFCM) estão um pouco mais misturados, não apresentando uma tendência geral bem definida. Uma explicação possível a esse resultado pode se encontrar na própria partição possibilística a qual permite que um mesmo documento possua um grau de tipicidade elevado em todos os grupos. Neste contexto, uma solução possível pode ser uma adaptação do método de extração de descritores SoftO-FDCL voltado para a partição possibilística, assim como também para algoritmos híbridos com duas partições, que é o caso do PFCM.

### 3.2 UMA ABORDAGEM HÍBRIDA PARA EXTRAÇÃO DE DESCRITORES

Nos experimentos anteriores foi identificado um possível problema ao realizar a extração dos descritores de maneira separada em cada partição do PFCM, assim como também foi apontado que o método pode não capturar toda essência da partição possibilística, que difere da partição fuzzy do FCM por não possuir a restrição que obriga a soma das pertinências de um grupo ser igual a um. Logo, é intuitivo indagar que para uma melhor interpretação dos grupos produzidos em um método de agrupamento híbrido, seja pertinente utilizar também uma abordagem mista de extração de descritores. Aproveitando-se assim dos benefícios existentes na partição possibilística, a qual penaliza os elementos ruidosos, com baixos valores de tipicidade, sem abrir mão das vantagens presentes na partição fuzzy. Para isso é necessário compreender os mecanismos de fun-



cionamento do método SoftO-FDCL, para que seja possível propor uma adaptação para este contexto.

### 3.2.1 INTERPRETANDO CORRETAMENTE AS TIPICIDADES

A interpretação direta dos graus de compatibilidade possibilísticos gera uma série de problemas na extração de descritores. Com isso, podemos formular a seguinte pergunta: Como interpretar corretamente os graus de compatibilidade possibilísticos para corretamente identificar os documentos relevantes de um dado grupo? Sabe-se que o valor de tipicidade pode variar livremente entre o intervalo [0, 1], sem a restrição probabilística do FCM. Essa é uma característica positiva introduzida em Krishnapuram e Keller (1993), a qual atribui valores de pertinência mais justos aos grupos fuzzy, em consonância com a teoria de conjuntos fuzzy, para melhor compreendermos esse conceito de valores mais justos.

Nesse contexto, propõe-se realizar tal interpretação em duas etapas. A primeira será constituída de uma conversão da tipicidade oriunda do PCM para a pertinência do FCM, de maneira a se satisfazer a restrição probabilística do FCM. No entanto, ao apenas realizar a conversão perde-se a robustez contra ruídos do PCM. Por isso, é possível contornar essa situação adicionando uma penalidade ao cálculo da pontuação dos termos.

A conversão proposta dos valores de tipicidade para pertinência, dar-se a como apresentado na Figura 5, a qual resolve o problema de considerar um documento como relevante em todos os grupos.

$$\lambda'(d_i, g_j) = \frac{\lambda(d_i, g_j)}{\sum_{k=1}^c \lambda(d_i, g_k)}$$

Figura 5

### 3.2.2 MÉTODO DE EXTRAÇÃO DE DESCRITORES PDCL

Agora que a proposta de interpretação da partição possibilística está concluída, o método para extração de descritores para a partição possibilística, o qual será denominado aqui de PDCL (Possibilistic Descriptor Comes Last) é proposto. Para realizar a extração de descritores, o método PDCL considera inicialmente todos os termos como candidatos. Em seguida, para cada grupo os valores de precisão e recuperação de todos os termos são calculados. A partir destes valores, a pontuação de cada termo no grupo é calculada, utilizando-se a medida  $f$ . A partir dessa pontuação por grupo dos termos candidatos, deve-se selecionar os descritores de maior pontuação em cada grupo. A quantidade de descritores é definida pelo usuário.

### 3.2.3 MÉTODO DE EXTRAÇÃO DE DESCRITORES Mixed-PFDCL

Uma das características presentes no método PFCM, é a adição dos parâmetros  $a$  e  $b$  que atuam como reguladores da influência do FCM e do PCM no agrupamento obtido. Portanto, é importante destacar a relevância de tais parâmetros no processo de extração de descritores, objetivando assim mais coerência com o algoritmo e exatidão nos resultados.

Nesse contexto, é também proposta a combinação dos valores de pertinência e de tipicidade convertendo-os em um único grau de compatibilidade conforme Figura 6. Essa combinação refere-se à média ponderada dos valores de pertinência e tipicidade pelos parâmetros  $a$  e  $b$  do método PFCM.

$$\mu'(d_i, g_j) = \frac{a\mu(d_i, g_j) + b\lambda'(d_i, g_j)}{a + b}$$

Figura 6



Como resultado dessa combinação, a estratégia definida anteriormente na Figura 4 é alterada, eliminando a dupla extração de descritores do método SoftO-FDCL do agrupamento produzido por meio do PFCM, na abordagem híbrida de extração de descritores com o método Mixed-PFDCL proposto. Essa nova estratégia proposta está contextualizada na Figura 7.

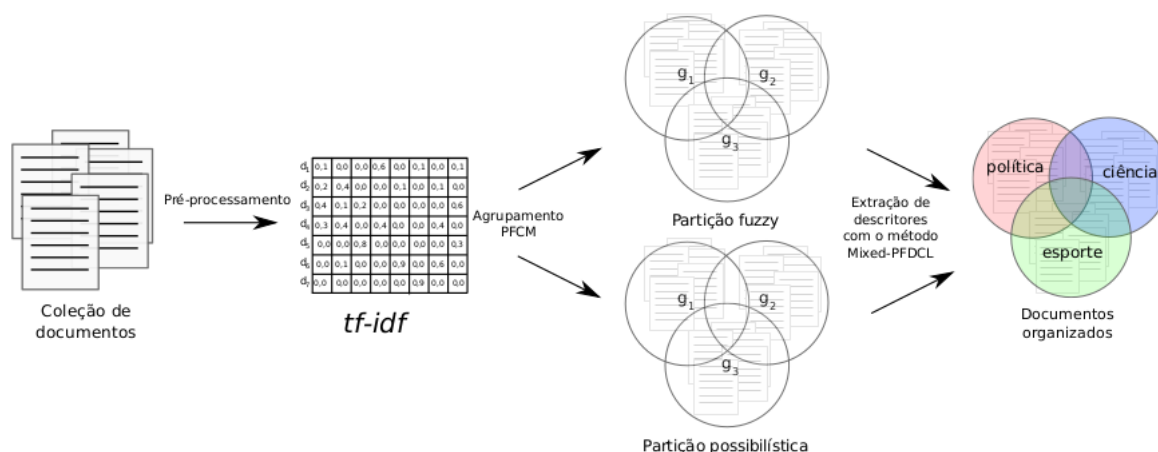


Figura 7: Estratégia híbrida proposta para uma organização flexível de documentos com o agrupamento com o método PFCM e a extração de descritores com o método Mixed-PFDCL.

Para mensurar os impactos das duas propostas (PDCL e Mixed-PFDCL) para extração de descritores apresentadas aqui nesta seção, foi realizado outro experimento, com os algoritmos PCM e PFCM. Durante o experimento foram utilizados os métodos de extração de descritores SoftO-FDCL, PDCL, Mixed-PFDCL.

Uma vez que as propostas aqui apresentadas têm por objetivo otimizar os resultados apresentados anteriormente, foi adotada uma metodologia similar aos experimentos anteriores. Ou seja, o agrupamento final obtido para cada base, é resultado dos grupos que obtiveram o maior valor na medida de silhueta fuzzy. Onde a quantidade de grupos para cada base, variou entre 2 e o número de classes de cada base (Tabela 7). Ressalta-se ainda que para minimizar os efeitos da aleatoriedade da partição inicial nesse experimento, o agrupamento foi executado 5 vezes para cada quantidade de grupos na silhueta fuzzy.

As coleções foram então agrupadas com os métodos PCM e PFCM, utilizando a metodologia descrita. A quantidade ótima de grupos, obtida com o método da silhueta fuzzy, está apresentado na Tabela 7. Os resultados apresentados nesta tabela, reforçam as conclusões apresentadas no experimento anterior, de que a quantidade de grupos ótima do método PFCM tende a se aproximar mais da quantidade original de classes de cada coleção, enquanto o método PCM possui uma tendência em obter um número de grupos bem inferior a quantidade original de classes.



Coleção	# classes	PCM	PFCM
Opinosis	3	2	3
20Newsgroup	4	4	4
Hitech	6	2	6
NSF	16	2	8
WAP	20	2	17
Reuters-21578	43	4	40

*Tabela 7: Quantidade ótima de grupos determinada através do método da silhueta fuzzy para cada algoritmo de agrupamento no segundo experimento conduzido com os métodos PCM e PFCM*

Em seguida, foi realizado a extração dos descritores sobre as partições ótimas encontradas por cada algoritmo de agrupamento, sobre as coleções textuais. Como a motivação desse experimento, foi avaliar a qualidade dos descritores produzidos pelos métodos PDCL e Mixed-PFDCL propostos, a extração de descritores foi também realizada com o método SoftO-FDCL, possibilitando assim compararmos os resultados. E assim como no experimento anterior, essa análise quantitativa dos descritores produzidos, foi feita, utilizando a mesma estratégia de avaliação preditiva, com os 5 algoritmos de classificação do experimento anterior.

O sumário desses resultados consta na Tabela 8, onde está marcado o método que obteve maiores taxas de acerto entre os 5 algoritmos de classificação utilizados. Como nessa investigação o propósito foi comparar os métodos de extração de descritores, dividiu-se o sumário de resultados de acordo com o método de agrupamento, PCM e PFCM respectivamente. Ressalta-se ainda, que garantir que a extração fosse realizada sobre os mesmos grupos, ambos os métodos de extração de descritores foram aplicados simultaneamente ao agrupamento. Portanto, os métodos SoftO-FDCL e PDCL foram aplicados ao mesmo agrupamento produzido pelo PCM, enquanto o SoftO-FDCL e o Mixed-PFDCL foram executados no mesmo agrupamento gerado pelo PFCM.

Os resultados dispostos nesse sumário, corroboram a hipótese formulada a respeito da interpretação das partições possibilísticas e híbridas no contexto da extração de descritores para a organização flexível de documentos. Pois, como se observa na Tabela 8, o método PDCL e o Mixed-PFDCL, superam os resultados do método SoftO-FDCL, em ambos os algoritmos de agrupamento. Embora, tenha existido 2 empates na comparação entre os métodos SoftO-FDCL e PDCL, para as coleções 20newsgroups e NSF.

Coleção	PCM		PFCM	
	SoftO-FDCL	PDCL	SoftO-FDCL	Mixed-PFDCL
Opinosis		✓		✓
20newsgroups	✓	✓		✓
Hitech		✓		✓
NSF	✓	✓		✓
WAP		✓		✓
Reuters-21578		✓		✓

*Tabela 8: Sumário dos resultados da classificação dos descritores extraídos com os métodos SoftO-FDCL, PDCL e Mixed-PFDCL*



#### 4. DISCUSSÃO

Ficou evidenciado, a partir das pesquisas encontradas na literatura, a diversidade de estratégias existentes na literatura para mitigar os desafios existentes. Foi visto ainda que tem sido frequentemente propostas abordagens para aprimorar todas as etapas existentes no processo de organização de documentos, as quais contemplam o pré-processamento, agrupamento, extração de descritores e recuperação da informação.

A partir do estudo da teoria e dos fundamentos necessários para o tema, foi apresentado o conteúdo abordando os detalhes dos métodos de agrupamento FCM, PCM, PFCM.

A primeira contribuição experimental dessa pesquisa, consistiu no estudo dos impactos de se adicionar o método PFCM no processo de organização flexível de documentos. A partir desse estudo, foi observado que o algoritmo PFCM possui uma tendência para aumentar a eficiência do agrupamento produzido em coleções textuais de maior dimensionalidade, o que foi comprovado a partir das evidências contidas na Tabela 4. No entanto, se destaca aqui a importância da realização de novos estudos com um maior número de coleções textuais de baixa e alta dimensionalidade para se confirmar esta tendência.

Por outro lado, constatou-se nesse experimento a capacidade de adaptação do método SoftO-FDCL a novos algoritmos de agrupamento. No entanto, este último método, segundo as suas pesquisas iniciais, considera somente uma única partição no processo de pontuação dos termos candidatos, o que por sua vez não consegue capturar toda a essência do agrupamento produzido pelo PFCM. Ainda nesse experimento foi observado um problema na interpretação das tipicidades contidas em partições possibilísticas, no processo de extração de descritores. Esse problema deriva diretamente da natureza probabilística dos graus de pertinência da partição fuzzy do FCM, que não se aplica às tipicidades, a qual influencia diretamente na adequação do limiar do método SoftO-FDCL.

Feitas essas considerações, apresentam-se as duas principais contribuições dessa pesquisa. A primeira consiste no método PDCL, que propõe uma abordagem para interpretar os graus de compatibilidade possibilísticos, sem deixar de lado a resiliência contra ruídos inerente das tipicidades. Os experimentos conduzidos com esse método demonstraram a qualidade dos descritores extraídos com o PDCL em comparação com o método SoftO-FDCL, cujos comparativos estão apresentados na Tabela 8. Também observou-se neste experimento que o método PCM produziu uma quantidade baixa de grupos em comparação ao número total de classes em cada coleção.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS (máximo 15)

[1] HAVENS, T. et al. Fuzzy c-means algorithms for very large data. IEEE Transactions on Fuzzy Systems, v. 20, n. 6, p. 1130-1146, 2012.

[2] HUANG, G. et al. A data as a product model for future consumption of big stream data in clouds. In: 2015 IEEE International Conference on Services Computing. Institute of Electrical & Electronics Engineers (IEEE), 2015. Disponível em: <http://dx.doi.org/10.1109/SCC.2015.43i>.

[3] MUGGLETON, S. H. 2020 computing: Exceeding human limits. Nature, Nature Publishing Group, v. 440, n. 7083, p. 409-410, mar 2006. Disponível em: <http://dx.doi.org/10.1038/440409ai>.

[4] KOBAYASHI, M.; AONO, M. Vector space models for search and cluster mining. In: Survey of Text Mining II. Springer Science + Business Media, 2008. p. 109-127. Disponível em: [http://dx.doi.org/10.1007/978-1-84800-046-9\\_6i](http://dx.doi.org/10.1007/978-1-84800-046-9_6i).



- [5] AGGARWAL, C. C.; ZHAI, C. An introduction to text mining. In: Mining Text Data. Springer Science + Business Media, 2012. p. 1-10. Disponível em: [http://dx.doi.org/10.1007/978-1-4614-3223-4\\_1i](http://dx.doi.org/10.1007/978-1-4614-3223-4_1i).
- [6] BAEZA-YATES, R.; RIBEIRO-NETO, B. Modern information retrieval: the concepts and technology behind search. Choice Reviews Online, American Library Association, v. 48, n. 12, p. 48-6950-48-6950, aug 2011. Disponível em: <http://dx.doi.org/10.5860/choice.48-6950i>.
- [7] NOGUEIRA, T. M. Organização Flexível de Documentos. Tese (Doutorado) — ICMC-USP, 2013.
- [8] BEZDEK, J. C.; EHRLICH, R.; FULL, W. Fcm: The fuzzy c-means clustering algorithm. Computers & Geosciences, v. 10, n. 2, p. 191 – 203, 1984. ISSN 0098-3004. Disponível em: <http://www.sciencedirect.com/science/article/pii/0098300484900207i>.
- [9] MACQUEEN, J. et al. Some methods for classification and analysis of multivariate observations. In: OAKLAND, CA, USA. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. [S.l.], 1967. v. 1, n. 14, p. 281-297.
- [10] KRISHNAPURAM, R.; KELLER, J. M. A possibilistic approach to clustering. IEEE Transactions on Fuzzy Systems, v. 1, n. 2, p. 98-110, 1993. ISSN 1063-6706.
- [11] PAL, N. R. et al. A possibilistic fuzzy c-means clustering algorithm. IEEE Transactions on Fuzzy Systems, IEEE Press, v. 13, n. 4, p. 517-530, 2005. ISSN 1063-6706.
- [12] HADDI, E.; LIU, X.; SHI, Y. The role of text pre-processing in sentiment analysis. Procedia Computer Science, v. 17, p. 26 – 32, 2013. ISSN 1877-0509. First International Conference on Information Technology and Quantitative Management. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1877050913001385i>.
- [13] FELDMAN, R.; SANGER, J. The text mining handbook: Advanced approaches in analyzing unstructured data. [S.l.]: Cambridge University Press, 2007.
- [14] ZHANG, C.; XU, H. Clustering description extraction based on statistical machine learning. Intelligent Information Technology Applications, 2007 Workshop on, IEEE Computer Society, Los Alamitos, CA, USA, v. 2, p. 22-26, 2008.
- [15] STEINBACH, M.; ERTÖZ, L.; KUMAR, V. The challenges of clustering high-dimensional data. In: In New Vistas in Statistical Physics: Applications in Econophysics, Bioinformatics, and Pattern Recognition. [S.l.]: Springer-Verlag, 2003. ISBN 978-3-642-07739-5.

## 6. ATIVIDADES REALIZADAS NO PERÍODO

- Leitura de artigos relacionados ao tema proposto.
- Desenvolvimento de métodos de agrupamento derivados do Fuzzy C Means.
- Desenvolvimento de métodos de extração de descritores.
- Implementação dos métodos de agrupamento clássicos da literatura.
- Realização de diversos experimentos comparativos.
- Escrita de artigo científico.
- Participação de reuniões do grupo de pesquisa.

## 7. PARTICIPAÇÃO EM REUNIÕES CIENTÍFICAS E PUBLICAÇÕES



Os resultados desta pesquisa foram detalhados no artigo científico em anexo, cuja submissão teve ótimas revisões e foi aprovado para publicação em uma das mais importantes conferências da área de sistemas fuzzy:

*Carvalho, N. V. J., Rezende S. O., Camargo H. A., Nogueira T. M. Flexible Document Organization by Mixing Fuzzy and Possibilistic Clustering algorithms. IEEE International Conference on Fuzzy Systems (FUZZ- IEEE), p. 1-8, 2016.*

## 8. DIFICULDADES ENCONTRADAS / CAUSAS E PROCEDIMENTOS PARA SUPERÁ-LAS

Por conta da excessiva carga de processamento requisitada nos experimentos, cada sequência de testes demandava demasiado tempo para conclusão, causando assim um atraso na obtenção dos resultados. Como solução de contorno foi alugado um serviço na nuvem para realizar o processamento. No entanto, ressalta-se que idealmente seria necessário o acesso a computadores com elevado potencial de processamento para a redução do tempo dos experimentos, possibilitando assim a realização de mais experimentos.

Salvador, \_\_\_\_\_ de \_\_\_\_\_ de 2016.

\_\_\_\_\_  
Estudante

\_\_\_\_\_  
Orientador (a)

Secretaria do Programa  
Rua Basílio da Gama, 06. Canela.  
Salvador – BA. 40.110-040.  
Tel.: 71 3283-7968 Fax: 71 3283-7964  
E-mail: [pibic@ufba.br](mailto:pibic@ufba.br)