

# COMPARATIVO ENTRE A TÉCNICA DE PADRÕES SEQUENCIAIS E CLUSTERIZAÇÃO PARA MINERAÇÃO DE CONTEÚDO APLICADA NA DETECÇÃO DE CONTEXTO EM AMBIENTE EDUCACIONAL

Patricia Mariotto Mozzaquatro Chicon - UNICRUZ - [patriciamozzaquatro@gmail.com](mailto:patriciamozzaquatro@gmail.com)  
Thiago Zanon Nunes - UNICRUZ - [zanonnunes@gmail.com](mailto:zanonnunes@gmail.com)

**Resumo.** Sistemas educacionais a distancia têm sido utilizados amplamente nos dias de hoje. A modalidade EAD a qual utiliza interface Web possui diversos recursos e ferramentas que são utilizados no processo de aprendizagem sendo considerada uma ótima alternativa ao usual método de ensino presencial. Entretanto, devido a infinidade de informações disponível na internet, muitas vezes torna-se uma árdua tarefa encontrar a mesma de forma precisa e confiável. Para tal problema destacam-se as técnicas de mineração de dados, mais especificamente a mineração web que se define como o desenvolvimento de técnicas e ferramentas para obtenção de conteúdo relevante útil na internet. No presente trabalho foram estudadas as técnicas de classificação e clusterização onde implementou-se os algoritmos de mineração de conteúdo Generalized Sequential Pattern algorithm (GSP) e Microsoft Clustering a fim de agrupar perfis semelhantes buscando similaridades entre eles, para assim, recomendar materiais educacionais.

**Palavras-chave:** *Mineração de dados Web. Algoritmo GSP. Algoritmo Microsoft Clustering.*

## COMPARATIVE BETWEEN THE TECHNICAL STANDARDS SEQUENTIAL AND CLUSTERING MINING APPLIED CONTENT IN DETECTION OF CONTEXT IN EDUCATIONAL ENVIRONMENT

**Abstract.** *Distance Education system have been widely used nowadays. The ODL mode which uses web interface has several features and tools that are used in the learning process is considered a great alternative to the usual method of teaching. However, due to the infinity of information available on the internet, often becomes and arduous task to find the same precise and reliable. For this issue we highlight the techniques of data mining, specifically the web mining which is defined as the development of techniques and tools for obtaining useful relevant content on the internet. In the presente work we study the techniques of classification and Clustering which was implemented algorithms content mining algorithm generalized sequential pattern (gsp) and Microsoft Clustering.*

**Keywords:** *Web mining, algorithm gsp, algorithm Microsoft Clustering*

### 1. INTRODUÇÃO

Devido à popularização da internet a área de educação na Web vem evoluindo nos últimos anos, de acordo com o relatório analítico da aprendizagem a distancia no Brasil publicado na página ABED – Associação Brasileira de educação a Distância ocorreu um crescimento significativo no número de matrículas realizadas no período de 2009 –

2011, no ano de 2009 ocorreram 528.320 matrículas contra 3.589.373 matrículas em 2011 (ABED, 2012).

No contexto educacional a maioria dos ambientes são páginas estáticas que apenas servem de repositórios de materiais, não oferecendo um serviço personalizado e interativo com o usuário. Neste sentido surge a necessidade de utilizar técnicas computacionais que detectem o contexto do usuário durante a interação com um ambiente educacional.

Como solução ao problema abordado pode-se utilizar técnicas computacionais que detectam o contexto do usuário, sugerindo ao mesmo, materiais educacionais personalizados e direcionados. Com um ambiente adaptado é possível proporcionar maior aproveitamento e desempenho dos acadêmicos que interagem com tais ambientes, ou seja, ocorre a filtragem de informações úteis na *web*, utilizando para isso técnicas de mineração *Web*, mais especificamente a mineração de conteúdo abordada no trabalho proposto.

A utilização de dois algoritmos *Generalized Sequential Pattern algorithm (GSP)* e *Microsoft Clustering*, integrantes da mineração de conteúdo, permitirá detectar o contexto do usuário em um ambiente educacional. O artigo proposto utilizou as técnicas padrões sequenciais e Clusterização a serem integradas a um ambiente educacional. Assim é possível descobrir semelhanças entre perfis de usuários e sugerir a eles conteúdos educacionais pertinentes a sua rotina de pesquisa e estudo, facilitando seu desenvolvimento intelectual e aprendizado.

## 2. MINERAÇÃO DE DADOS WEB

A rede mundial de computadores é um conjunto de computadores e servidores interligados com um amplo repositório de dados, que atualmente se tornou uma ferramenta muito utilizada para a disseminação de conhecimento e troca de informação (BOHN, 2010). Neste contexto, se torna complicado encontrar informação correta devido ao seu exponencial crescimento, por isso se faz necessário realizar um processo de filtragem dos dados adquiridos com pesquisas na rede. Assim destaca-se o emprego de técnicas de mineração de dados *web*.

De acordo com (MAGALHÃES, 2008):

A mineração *Web* é uma área de pesquisa que visa integrar as tecnologias *Web* e a mineração de dados, focalizando o desenvolvimento de novas ferramentas e métodos para análise e descoberta de conhecimento de dados na *web*. Ela pode ser definida como a descoberta e análise de informações úteis, novas e interessantes na internet, onde a partir das informações descobertas, seja possível demonstrar características, comportamentos, tendências e padrões de navegação do usuário na *web*.

Entretanto, minerar os dados encontrados em páginas *web* trata-se de uma tarefa mais complicada com relação à mineração de dados tradicional, pois as páginas da *web* não se encontram em uma estrutura organizada de quadros e colunas como consiste um banco de dados, os dados na internet estão agrupados em uma estrutura de *hiperlinks* muitas vezes interligados ou não, sendo que a estrutura das páginas não contém nenhuma organização comum entre elas, com isso não existe nenhum controle sobre os documentos que são postados na rede.

A mineração *Web* pode ser dividida em três subáreas sendo elas, mineração de uso, de conteúdo e de estrutura. A pesquisa desenvolvida integrou a Mineração de conteúdo. A mineração de conteúdo ou *web content mining* é um segmento de pesquisa que visa extrair informações de páginas da internet de acordo com seu conteúdo, suas aplicações são utilizadas em diversos tipos de dados como texto, imagem, áudio, vídeo e meta dados (SOUZA, 2009).

Ao aplicar a mineração de conteúdo encontra-se grande dificuldade para adquirir informações, pois a maioria dos documentos na internet não obedece a nenhuma normatização em sua estrutura, entretanto com o passar do tempo foram desenvolvidas técnicas computacionais com o objetivo de conseguir uma maior estruturação dos documentos na web, tais técnicas buscam extrair informações relativas as características do domínio e perfis dos usuários de forma a organizar e interpretar as informações adquiridas (BOULLOSA, 2002). As seções a seguir irão apresentar as técnica, juntamente com seus algoritmos de padrões sequenciais e clusterização.

### 3. TÉCNICA DE PADRÕES SEQUENCIAIS

Segundo (Barbosa, 2006) padrões sequenciais são eventos que ocorrem no tempo e que possibilitam ser utilizado para prever eventos futuros baseados nos anteriores, um exemplo de padrão sequencial é uma pessoa que adquire um celular hoje, é um potencial comprador daqui a alguns meses.

Os padrões sequenciais buscam localizar relação entre eventos que ocorrem em sequencias a fim de determinar uma ordem nessas ocorrências, é possível localizar padrões de sequenciais de um item específico e também é possível localizar padrões cruzando informações de diferentes itens, mineração de padrões sequenciais é muito utilizado para análises de sequencias de DNA (ZHAO, 2003). A subseção a seguir apresenta o Algoritmo *Generalized Sequential Pattern* (GSP).

#### 3.1 Algoritmo *Generalized Sequential Pattern* (GSP)

GSP é um algoritmo concebido para minerar padrões sequenciais, seu funcionamento acontece iterativamente onde em cada interação a base de dados é totalmente percorrida a fim de encontrar todas as sequencias ali contidas, onde cada interação é denotada por  $k$ , ou seja, sequencias encontrada na base de dados possuem tamanho  $k$  (BARBOSA, 2006).

O algoritmo efetua várias verificações na base de dados, onde primeiramente é determinada a quantidade de sequencias de dados que será incluído em cada item, ou seja, é determinado o suporte de cada item, ao final dessa sequencia o algoritmo é capaz de determinar os itens que cumprem os parâmetros mínimos especificados de suporte mínimo, onde cada item é uma sequencia de elementos frequentes que possui tamanho 1 (CALAPEZ, 2008).

O algoritmo GSP é baseado no Apriori, porém se difere em alguns aspectos, no GSP, a cada interação  $k$  os conjuntos  $L_k$  (*Itemsets* frequentes) e  $C_k$  (*itemsets* candidatos) são formados por sequencias de  $k$  *itemsets*, o processo que o algoritmo GSP executa pode ser resumido em, geração de candidatos, calculo de suporte e poda de candidatos (DEVEZA, 2011).

## 4. TÉCNICA DE CLUSTERIZAÇÃO

As técnicas de Clusterização ou agrupamento são técnicas onde o cálculo é efetuado no que se refere ao número de grupos ou sua estrutura, o agrupamento é efetuado a partir das distâncias ou similaridades entre os componentes, onde as únicas condições são medidas de similaridade ou dados sob os quais possam ser calculadas as similaridades (KASZNAR, 2007).

O objetivo dos algoritmos de Clusterização é o agrupamento de padrões em conjuntos de dados, com a finalidade de detectar igualdade entre objetos e determinar uma classificação de acordo com suas propriedades em comum, os algoritmos de Clusterização criam segmentos chamados de clusters onde um cluster é um conjunto ou grupo de dados aos quais os mesmos são formados de acordo com similaridades (GOLE, 2001). A subseção a seguir irá apresentar o algoritmo *Microsoft Clustering*.

### 4.1 Algoritmo *Microsoft Clustering*

O algoritmo *Microsoft Clustering* é um algoritmo que analisa e divide os dados em grupos de acordo com a semelhança entre os dados, utilizando um conjunto de atributos e parâmetros, após a definição dos agrupamentos os mesmos são definidos evidenciando os valores que melhor os representam (MENDES, et, al, 2008).

O algoritmo *Microsoft Clustering* é uma ferramenta oferecida como recurso do *analysis services*, utiliza *clusters* agrupando dados de acordo com suas similaridades, sendo utilizados para detectar anomalias nos dados e criar previsões, tais modelos detectam características e relações nos dados, algo que não pode ser realizado por deduções e observações de pessoas, por exemplo, uma pessoa é capaz de determinar que pessoas que utilizam bicicletas para irem ao trabalho não moram longe do mesmo, porém o algoritmo é capaz de encontrar características não óbvias.

O funcionamento do algoritmo *Microsoft Clustering* é realizado da seguinte maneira: determina a relação de acordo com similaridades e define os primeiros clusters, após a definição dos primeiros clusters o algoritmo efetua cálculo para determinar o agrupamento dos pontos, a partir desse ponto o algoritmo realiza uma redefinição nos agrupamentos para criação de novos clusters que definirão melhor a representação dos dados, o algoritmo efetua essa redefinição até que não seja mais possível melhorar o resultado pelas redefinições dos clusters (MICROSOFT, 2013).

O *Microsoft Clustering* é um algoritmo que faz parte do conjunto de ferramentas do *analyses services*, que por sua vez é utilizado dentro do sistema gerenciador de banco de dados *Microsoft SQL Server*, esse sistema permite que seja apresentado a representação dos dados na forma de rede, não só na forma tabular como ocorre normalmente em consultas, mas também com arcos e uma codificação em cores, algo que torna claro as relações dos dados nos *clusters* (ÁLFARO, et, al, 2008).

## 5. SISTEMA DESENVOLVIDO

O sistema desenvolvido tem como objetivo proporcionar ao usuário um maior aproveitamento e conseqüentemente desempenho em aula do acadêmico, pois irá possibilitar a indicação de materiais de estudo de forma confiável de acordo com seu interesse.

A Figura 1 ilustra o diagrama de caso de uso do sistema proposto. No referido diagrama o usuário acessa o sistema, o mesmo realiza autenticação, se já possui cadastro

é direcionado a próxima página do sistema, caso contrário, o sistema emite uma mensagem informando ao usuário um erro.

Permitido o acesso ao ambiente o usuário poderá efetuar a navegação nos *links* das páginas onde se encontram os materiais postados pelos professores. Conforme a interação do usuário com o sistema, o mesmo irá captar suas atividades e por meio dos cliques do *mouse* e *logs* de acesso será possível gerar o perfil desses usuários.

Com a aplicação dos algoritmos *GSP* e *Microsoft Clustering* será gerado o perfil dos usuários. Após a geração dos perfis o sistema irá sugerir materiais e atividades, gerando assim, grupos de usuários de acordo com o perfil que foi gerado pelos algoritmos.

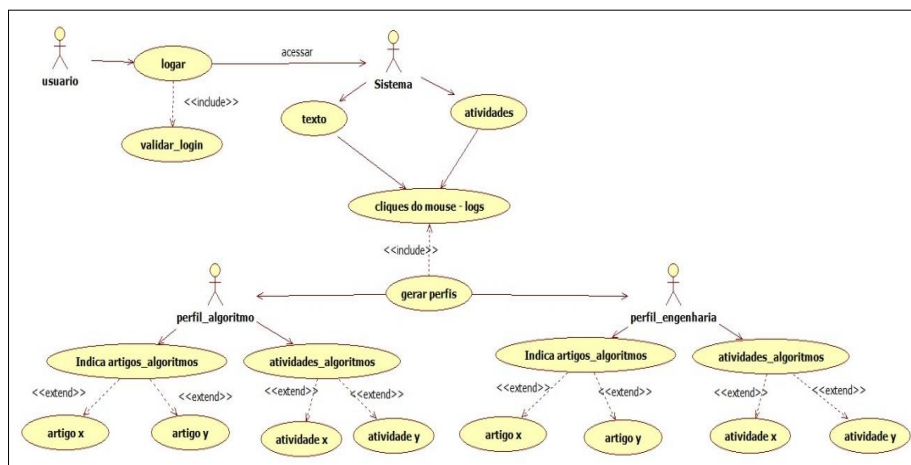


Figura 1 – Diagrama de casos de uso

As subseções a seguir irão apresentar a integração dos algoritmos *GSP* e *Microsoft clustering* ao sistema proposto.

### 5.1 Integração do algoritmo *GSP*

A técnica do algoritmo *GSP* foi implementada dentro do sistema desenvolvido, para que o algoritmo possa efetuar sua verificação, o mesmo necessita de alguns dados como pré-requisitos para efetuar seu rastreamento e buscar determinar perfis de usuários, esses pré-requisitos são tempo, cliques dos usuários, e determinação de suporte mínimo do algoritmo.

Após o usuário acessar o sistema o cronometro é disparado, no referido algoritmo. Foi determinado dez minutos para as primeiras interações com o ambiente, com isso, a tabela registro foi previamente alimentada com o tempo e o suporte mínimo. Para suporte mínimo inicial foi determinado cinco (Barbosa, 2006), ao acessar os *links* do ambiente a tabela contador vai sendo alimentada com os cliques do *mouse* no momento da interação com o ambiente.

Se ao término dos dez minutos o usuário tiver efetuado a quantidade de cliques em uma mesma categoria disponível do sistema, o algoritmo efetua uma comparação com o suporte mínimo.

Caso a quantidade de cliques tiver alcançado ou ultrapassado esse parâmetro nos primeiros dez minutos o perfil é detectado. Caso contrário, os dados da tabela registro são atualizados e o cronometro é incrementado com mais cinco minutos, o suporte mínimo também é incrementado com um, ou seja, passou de cinco para seis, esse

incremento é efetuado até que as condições para detecção do perfil pelo algoritmo sejam satisfatórias.

Caso essas verificações tragam resultados satisfatórios para a detecção do perfil pelo algoritmo, o sistema emite mensagem para o aluno informando a detecção do perfil, e redireciona o mesmo para outra página, assim o algoritmo termina sua verificação. Caso esta condição não seja satisfeita, o mesmo é redirecionado novamente até a página inicial e o cronometro é reiniciado, esse ciclo ocorre recursivamente até que o algoritmo seja capaz de realizar a detecção do perfil do usuário.

## 5.2 Integração do algoritmo *Microsoft Clustering*

A implementação do algoritmo *Microsoft Clustering* ocorre de maneira similar a do *GSP*, porém o mesmo necessita de menos verificações e parâmetros para a geração dos *clusters* e detecção do perfil do usuário.

Primeiramente, ao usuário acessar o sistema o mesmo realiza uma inserção de dados na tabela registro, nessa tabela são salvas informações de tempo inicial, tempo final e identificador do usuário. Após essa inserção no primeiro acesso do sistema, o usuário é direcionado até a página inicial para as primeiras interações, no referido trabalho o tempo de contagem regressiva do cronometro foi de cinco minutos.

Ao interagir com o ambiente os dados de cliques do *mouse* são salvos na tabela contador, juntamente com o identificador único de cada usuário e com o identificador de cada categoria clicada, esse processo é realizado até o cronometro chegar a zero, após chegar a zero o sistema realiza verificação para a geração dos *clusters* e detecção do perfil, caso os pré-requisitos para essa verificação estejam corretos.

Como no algoritmo *Microsoft Clustering* não existe parâmetros como suporte mínimo, suas verificações para detecção do perfil são realizadas de maneira geral de modo mais simples.

Ao final de todas as verificações, quando os algoritmos tiverem detectado o perfil dos usuários, o sistema será capaz de sugerir os materiais direcionados de acordo com o contexto de cada aluno, para isso o sistema deve sofrer uma adaptação em sua interface, para que o aluno possa acessar esses materiais.

Assim quando o sistema enviar mensagem de aviso ao aluno informando que seu perfil foi detectado, o mesmo irá direcioná-lo a outra página com um *link* no canto superior direito com as respectivas sugestões de materiais.

## 6. RESULTADOS E DISCUSSÕES

No presente estudo foram aplicados diversos testes com a intenção de encontrar possíveis falhas e erros no sistema desenvolvido, os testes aplicados foram caixa branca e caixa preta.

O teste caixa branca consiste em uma análise realizada no código fonte do sistema e na sua lógica de desenvolvimento e funcionamento, é testada toda a parte do *back-end* do sistema, nesse tipo de teste é realizada uma busca por erros na lógica e no comportamento do sistema. Os testes de caixa branca utilizados foram teste de unidade, validação e segurança (BEQUE, 2009).

O teste de unidade busca identificar erros na lógica do *software* separadamente em cada modulo do sistema (Fontoura, 2008), assim foram testadas de maneira separadas as funcionalidades de cada módulo, no módulo do administrador foi testado o cadastro de usuários, cadastro de materiais auxiliares e toda a parte de geração de

relatórios e comparação de dados oriundos dos relatórios e os dados contidos no banco de dados.

O teste de validação verifica se as funções criadas no sistema estão funcionando de maneira correta, ou seja, se cada função criada está realizando suas atribuições de acordo com as especificações escritas nos códigos do sistema. (SOMMERVILLE, 2011).

Foi realizado teste de segurança com relação ao acesso dos usuários, para isso foi feita diversas tentativas de acesso com usuários sem cadastro no sistema, só é permitida entrada de usuários previamente cadastrados pelo usuário administrador.

Também existe um controle de seção para cada usuário, assim se algum usuário tentar acessar o sistema digitando diretamente o endereço de uma das páginas internas sem efetuar *login* o mesmo irá emitir uma mensagem de erro informando que para acessar essa página deve-se ter efetuado *login* no sistema.

Os testes caixa preta são testes efetuados com relação ao *front-end* do *software*, ou seja, toda a parte de interface e interação com os usuários, os usuários tem a informação do que se espera do sistema, o que ele devera realizar sem ter conhecimento sobre a estrutura dos códigos e a lógica aplicada para o desenvolvimento do mesmo.

O teste caixa preta foi aplicado a uma população de 35 alunos, sendo que metade dessa população está entre o sexto e sétimo semestre, o restante se encontra no segundo semestre do curso de ciência da computação da Universidade de Cruz Alta.

Conforme o autor (Barbetta, 2002) o teste aplicado a uma população finita deve integrar, no caso do trabalho proposto, alunos de semestre inicial e final. Primeiramente foi explicado aos alunos o objetivo da pesquisa e realizada uma introdução sobre o sistema proposto.

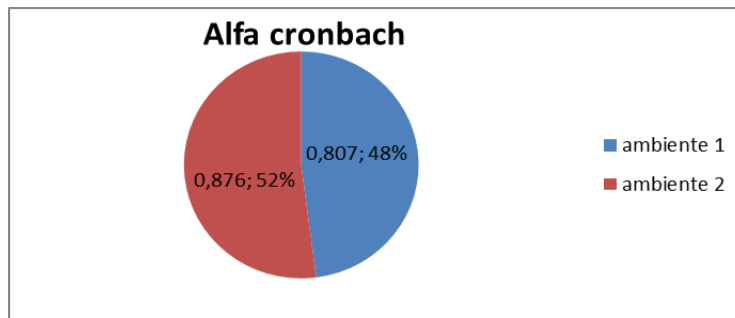
Após, foi efetuado testes do algoritmo *GSP* onde foi estipulado primeiramente tempo de dez minutos para interação dos usuários no sistema. Após dez minutos foi incrementado mais cinco para assim finalizar a varredura do algoritmo nos dados coletados na navegação desses usuários.

No segundo momento foi efetuado teste do algoritmo *Microsoft Clustering* onde foi determinado tempo de cinco minutos para a detecção do perfil dos usuários. Quando finalizou a interação com o sistema foram aplicados dois questionários, um para cada algoritmo com onze perguntas cada relacionadas aos seguintes domínios: interação, recomendação e usabilidade do sistema.

Para a elaboração do questionário foi utilizado o *software makesurvey*, essa ferramenta *web* disponibiliza uma interface de criação, edição e geração de relatórios das perguntas dos dois questionários implementados.

Para a validação das respostas dos usuários foi utilizado o *software IBM SPSS statistics* versão 2.0, utilizando a métrica Alfa de *Crombach* para análise dos dados coletados. O teste Alfa de *Crombach* é um medidor de confiabilidade, controla o grau de covariância de um número determinado de itens que varia de 0 a 1, quanto mais elevada a contagem, maior é a confiabilidade dos resultados obtidos (ARAÚJO, 2009).

O coeficiente de Alfa de *Crombach* foi verificado de acordo com (Almeida, 2010) o resultado mínimo esperado para validação do alfa é 0,70, abaixo desse valor a consistência é conceituada como baixa, o valor máximo esperado é 0,90, acima desse valor pode ser considerado redundância ou duplicação. O gráfico 1 demonstra os valores obtidos nos dois ambientes aplicando o coeficiente Alfa de *Crombach*.

Gráfico 1 - Resultados coeficiente *alfa cronbach*

O Gráfico 1 ilustra os resultados obtidos do coeficiente Alfa de *Cronbach*, é possível verificar que o algoritmo *Microsoft Clustering* obteve um valor maior de coeficiente, ou seja, o ambiente dois possui uma consistência interna maior com relação ao algoritmo *GSP*.

Também foi verificado o desvio padrão na análise dos resultados. O desvio padrão verifica o quanto foram dispersos os resultados das respostas dos alunos no questionário, ou seja, quanto maior o desvio padrão maior é a dispersão de respostas com relação a média (Gráfico 2).

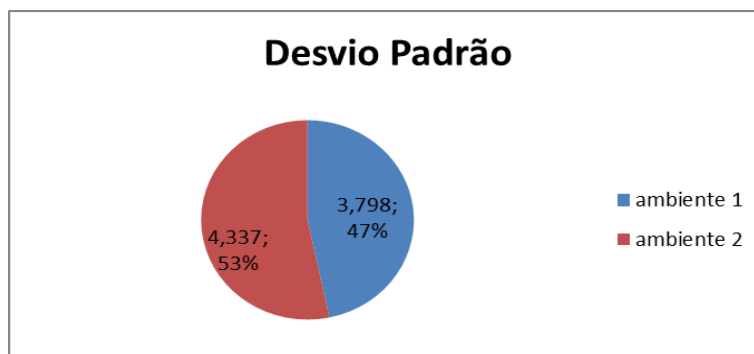


Gráfico 2 - Desvio padrão

No Gráfico 2 é possível verificar que o desvio padrão do algoritmo *Microsoft Clustering* obteve um valor maior com relação ao algoritmo *Gsp*, ou seja, o ambiente dois demonstrou uma média de dispersão maior entre as respostas, evidenciando uma maior variedade nas respostas dos usuários.

O Gráfico 3 exibe os resultados das médias alcançadas em cada um dos ambientes implementados.

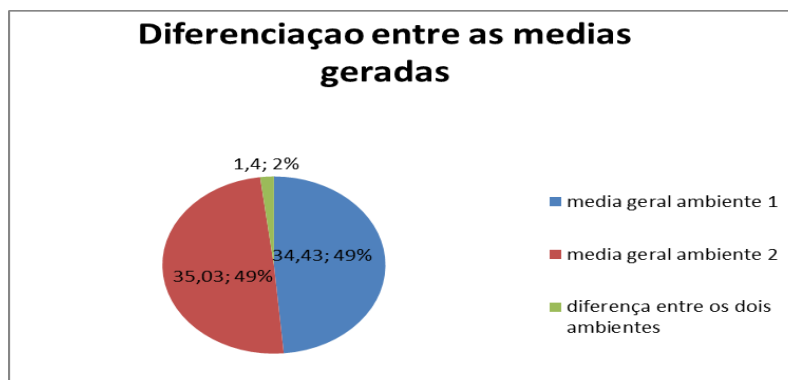


Gráfico 3 - Média dos ambientes



O gráfico 3 mostra os valores da média geral e a diferença apresentada entre os dois ambientes implementados. O algoritmo *Gsp* alcançou 34,43% e o algoritmo *Microsoft Clustering* 35,03%, assim é possível verificar que o ambiente dois alcançou uma média maior com relação ao ambiente um, apesar da pouca diferença entre os resultados apresentados na validação dos dois algoritmos. Após o estudo desenvolvido fica evidenciado que no contexto geral o algoritmo *Microsoft Clustering* obteve resultados mais contundentes e satisfatórios que o algoritmo *GSP*.

## 7. CONSIDERAÇÕES FINAIS

A utilização de recursos computacionais para desempenhar atividades e auxiliar de maneira correta a execução das mais diversas tarefas é algo de muita importância nos dias de hoje.

O presente estudo teve como objetivo comparar dois algoritmos de Mineração de conteúdo da *web* *Generalized Sequential Pattern algorithm* e algoritmo *Microsoft Clustering*, medindo sua eficiência na detecção de contexto do usuário durante a interação com um ambiente educacional a fim de agrupar perfis semelhantes, buscando similaridades entre eles e recomendando materiais educacionais.

Para a aplicação dos dois algoritmos foi desenvolvido um sistema que simula um ambiente educacional utilizando a linguagem de programação *php* juntamente com o *SGBD mysql*, ou seja as duas técnicas tiveram validações em um ambiente *web*.

A partir das implementações dos algoritmos no ambiente, nos testes executados de caixa branca e preta, e dos resultados obtidos por meio da análise da métrica alfa de *cronbach*, pode-se concluir que o algoritmo *Microsoft Clustering* obteve melhor desempenho na tarefa proposta para esse trabalho.

O resultado obtido justifica-se além da análise de resultados que o mesmo necessita de menos parâmetros para execução de suas tarefas e conseqüentemente detecção do perfil do usuário no sistema.

Foi evidenciado a partir da implementação do algoritmo *GSP* no ambiente os três problemas indicados por (Cavique, 2008) na primeira parte do desenvolvimento do trabalho, segundo o referido as técnicas de padrões sequenciais possuem três problemas, precisa de diversos parâmetros, escalabilidade dos algoritmos e a limitação global dos padrões, nos parâmetros o utilizador deve determinar um limite de suporte mínimo para localizar os padrões, se o utilizador determinar um valor de suporte mínimo muito alto pode ocorrer corte excessivo ou se for muito baixo pode ocorrer um número reduzido de itens.

A partir dessas evidências é possível determinar que o presente estudo contribui cientificamente, pois provou que o algoritmo *Microsoft Clustering* teve melhor desempenho que o algoritmo *GSP* na execução dessa tarefa, evidenciando uma tendência computacional que é de descobrir o que um usuário busca dentro de um determinado ambiente e auxiliá-lo para um melhor aproveitamento na descoberta do saber e do conhecimento.

O presente estudo visou testar a eficiência das duas técnicas para a detecção de contexto de determinado usuário no ambiente, assim com o êxito dessa aplicação é possível determinar que com um perfil gerado é muito mais fácil ajudar um aluno em estágio de aprendizado dentro da *internet*, e isso é mais uma opção a todos os métodos de ensino conhecidos hoje.

Desse modo, a Mineração de dados aplicada na área da educação, conforme o artigo apresentado, contribui cientificamente em pesquisas futuras, pois no Brasil é um campo do conhecimento pouco difundido, atualmente não se tem muitas pesquisas nesta área (BAKER, 2010).

## REFERÊNCIAS BIBLIOGRÁFICAS

ABED - **Associação Brasileira de educação a Distância**. Disponível em;<<http://www.abed.org.br/>>.Acessado em: maio. 2015.

ÁLFARO, Paulo Jorge, MENDES Armando, FERREIRA Aires. **Suporte à Decisão em Tecnologias de comunicação: Um Caso Com Utilização de Metodologias OLAP e Data Mining**. Centro de Estudos de Economia Aplicada do Atlântico – CEEAPLA WP No. 09/2008.

BEQUE, Luciéli Tolfo. **Avaliação dos Requisitos para Teste de Um sistema operacional Embarcado**. Dissertação de mestrado do programa de pós-graduação em computação da Universidade Federal do Rio Grande do sul – UFRGS, 2009.

BAKER, R.S.J.D. **Data Mining for Education**. McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education (3rd edition)*. Oxford, UK: Elsevier. 2010.

BARBOSA, Ciro Bastos, PONTE, Eduardo, PLASTINO, Alexandre; **Explorando Técnicas de Redução de Base de Dados na Mineração de Padrões Sequenciais**. Dissertação de Mestrado do programa de pós-graduação de computação da Universidade Federal Fluminense – UFF, 2006.

CAVIQUE, Luís, COELHO, José Silva; **Descoberta de Padrões Sequenciais Utilizando Árvores Orientadas**. Revista de Ciências da Computação, volume III – 2008.

CALAPEZ, Marco Miguel Guerreiro; **SISMAC: Sistema de Monitoria de Alterações Climáticas**. Dissertação de Mestrado em Engenharia Informática e de Computadores da Universidade Técnica de Lisboa – UTL, 2008.

DEVÊZA, Cecília Henriques; **Mineração de dados para padrões de sequência**. Monografia Apresentada ao curso de Bacharelado em Ciência da Computação Universidade Federal de Ouro Preto – UFOP, 2010.

DEVÊZA, Cecília Henriques, **Minerando Padrões Sequenciais para Base de dados de Lojas Virtuais**. Monografia apresentada ao curso de ciência da computação da Universidade Federal de Ouro Preto – UFOP, 2011.

METZ, Jean; **Interpretação de Clusters gerados por algoritmos de Clustering Hierárquico**. Dissertação de Mestrado apresentada ao Instituto de ciências matemáticas e de computação – ICMC – USP, 2006.

MICROSOFT - Disponível em; <<http://msdn.microsoft.com>> Acessado em: Junho 2013.

SOMMERVILLE, Ian, **Engenharia de Software**. 9º edição, p.540, 2011.