UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Algoritmos de Extração de Palavras-Chave de Textos em Português

Marcel Brito Pereira
Maria das Graças Volpe Nunes

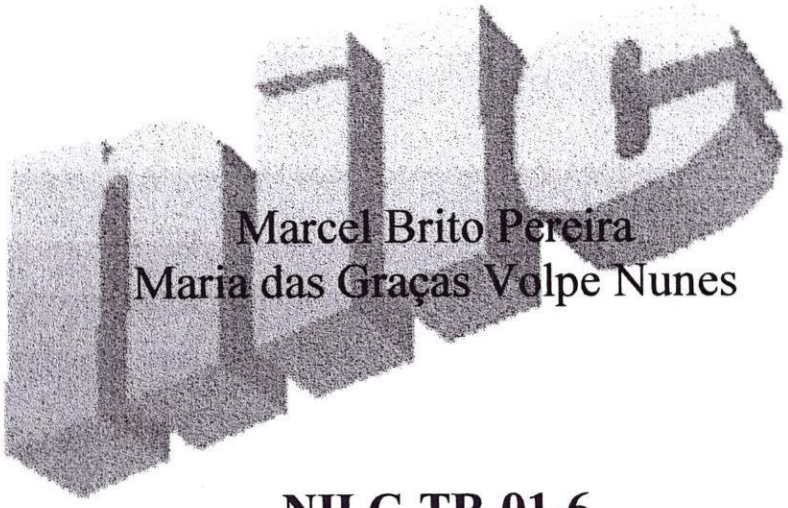
Nº 151

RELATÓRIOS TÉCNICOS

São Carlos - SP

Universidade de São Paulo - USP
Universidade Federal de São Carlos - UFSCar
Universidade Estadual Paulista - UNESP

ALGORITMOS DE EXTRAÇÃO DE PALAVRAS-CHAVE DE TEXTOS EM PORTUGUÊS



Marcel Brito Pereira
Maria das Graças Volpe Nunes

NILC-TR-01-6

Outubro, 2001

Série de Relatórios do Núcleo Interinstitucional de Lingüística Computacional
NILC - ICMC-USP, Caixa Postal 668, 13560-970 São Carlos, SP, Brasil

SYSNO 1211506
DATA / /
ICMC - SBAB

ALGORITMOS DE EXTRAÇÃO DE PALAVRAS-CHAVE DE TEXTOS EM PORTUGUÊS*

Marcel Brito Pereira
marcelbp@grad.icmc.sc.usp.br
Maria das Graças Volpe Nunes
mdgvnune@icmc.sc.usp.br
NILC - Núcleo Interinstitucional de Lingüística Computacional
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo - São Carlos

RESUMO

Algoritmos de extração de palavras-chave de textos podem ser bastante úteis em diversas aplicações. Palavras-chave constituem um mecanismo clássico de indexação de documentos e, em aplicações na web, têm se tornado cada vez mais relevantes. Este trabalho apresenta o projeto e a implementação de dois algoritmos de extração de palavras-chave de textos em português, baseados em métodos extrativos, comparando-os e analisando seu desempenho.

ABSTRACT

Keywords have especial role in the process of indexing documents. Many web applications can take advantage of good extraction algorithms. This work presents and discusses the performance of two keyword extraction algorithms for Portuguese.

1 INTRODUÇÃO

Palavras-chave (keywords ou keyphrases) são muito úteis em aplicações que necessitem indexar documentos para buscas posteriores. Para a língua inglesa, muitos trabalhos apresentam técnicas diversas de extração de palavras-chave de textos. Para o português, no entanto, desconhecem-se técnicas apresentadas na literatura especializada. O trabalho apresentado aqui tem o objetivo de investigar o desempenho de dois algoritmos de extração de palavras-chave de textos em português, usando técnicas extrativas simples, porém recursos metodológicos que estão de acordo com os trabalhos desenvolvidos para outras línguas.

A principal aplicação que motivou esse estudo - embora não a única - é a geração automática de sumários (resumos) extrativos de textos em português. A sumarização extrativa, ao contrário da fundamentalista, emprega técnicas puramente estatísticas e superficiais para a seleção de um subconjunto de sentenças do texto original que possam expressar a idéia central do mesmo (Black & Johnson, 1998). Enquanto que os métodos fundamentais se baseiam em técnicas sofisticadas de processamento de línguas naturais, pagando um alto preço de projeto, as extrativas são fortemente baseadas na presença de palavras-chave e algumas outras "pistas"

* Trabalho de IC com apoio do PIBIC/CNPq.

simples para a seleção de informação relevante. Evidentemente, essa simplicidade paga seu preço com a falta de coesão e coerência do texto gerado (Sparck Jones, 1993). Dependendo da aplicação a que se destina, no entanto, esse preço pode ser compensador. Tendo isto em vista, os algoritmos desenvolvidos nesse trabalho foram utilizados para a criação do SUMEX (Souza, 2001), programa este que visa a geração automática de sumários extrativos de textos em português.

Este artigo relata o projeto, a implementação, a comparação e a análise de desempenho de dois algoritmos de extração de palavras-chave de textos científicos em português. O primeiro, EPC-P, é baseado na freqüência de determinados padrões morfossintáticos obtidos de um corpus de artigos científicos, conforme discutido na Seção 3. O segundo, EPC-R, é uma adaptação do algoritmo Extractor (Turney, 1997;1999), para o inglês, e se baseia nas freqüências de radicais, e será discutido na Seção 4.

Na Seção 2 é apresentada uma metodologia de avaliação de algoritmos de extração de palavras-chave, que será utilizada para a avaliação dos algoritmos implementados. Na Seção 5 são apresentados e analisados os resultados obtidos, sendo o trabalho concluído na Seção 6. O Apêndice I traz os resultados de alguns dos testes realizados.

2 MEDINDO A PERFORMANCE DE UM ALGORITMO DE PALAVRAS-CHAVE

A avaliação de um algoritmo de extração de palavras-chave pode ser muito subjetiva quando feita a partir da leitura e análise do texto original, quer seja pelo autor ou por terceiros. No entanto, qualquer método alternativo a esse corre o risco de ignorar critérios de relevância altamente dependentes de significado e intenções, portanto, difíceis de se obter através de técnicas extrativas simples. Para minimizar esse gargalo de avaliação, optamos por aplicar os algoritmos a textos científicos para os quais os autores forneceram um conjunto pequeno de palavras-chave, comparando-as com o resultado dos algoritmos. Evidentemente, o conjunto do autor é muito reduzido, se considerarmos aplicações em que o limite usualmente imposto pelas publicações não seria relevante. Dessa forma, o conjunto produzido pelo algoritmo será substancialmente maior do que o do autor e a indicação de uma palavra-chave que não ocorre na lista do autor não significa necessariamente um mau desempenho do algoritmo. Já a ausência de uma palavra-chave do conjunto do autor no conjunto produzido pelo algoritmo poderá indicar deficiência, desde que não ocorram outras similares. Acrescenta-se a esse cenário o fato de que nem sempre as palavras-chave do autor expressam corretamente os tópicos compreendidos pelo artigo. É comum que uma certa porcentagem das palavras-chave elaboradas pelos autores dos artigos sequer apareça no texto, ou apareça com baixa freqüência, fazendo com que métodos estatísticos baseados em freqüência de ocorrência não consigam selecioná-las. Isso pode ocorrer por várias razões: (1) o autor pode preferir adotar como palavra-chave um termo mais geral; (2) num artigo científico em português, ele opta usar como palavra-chave o termo mais conhecido em inglês, e no texto usa sua tradução; (3) o autor escolhe as palavras-chave de forma usualmente subjetiva e aleatória.

Mesmo sendo pouco determinísticos, no entanto, é desejável que esses critérios se traduzam em valores numéricos, para uma fácil comparação. A maioria dos trabalhos similares em recuperação da informação usa as medidas complementares de *precisão (precision)* e *recuperação (recall)* para medir a performance de seus algoritmos. A medida de precisão fornece a probabilidade de que uma dada palavra que o programa de extração classificou como palavra-chave também tenha sido classificada como tal pelo autor do artigo. A de recuperação fornece a probabilidade de que uma dada palavra classificada como palavra-chave pelo autor também seja classificada como tal pelo programa.

3 EPC-P - EXTRAÇÃO DE PALAVRAS-CHAVE BASEADA EM FREQÜÊNCIA DE PADRÕES

Neste primeiro método, que chamamos simplesmente de EPC-P (Extrator de Palavras-Chave por freqüência de Padrões), decidimos por construir um sistema que busque no texto conjuntos de palavras que se encaixem em determinados padrões morfossintáticos. Depois de realizado o levantamento de todas as frases que se encaixam nos padrões, utilizamo-nos de métodos estatísticos para definir, dentre estas, quais seriam as de maior relevância, podendo ser consideradas palavras-chave para o texto.

A seguir, apresentamos em detalhes a fase de levantamento dos padrões, assim como uma descrição verbal do algoritmo.

3.1 Levantamento de Padrões

Foi realizado um levantamento em 12 exemplares de revistas científicas brasileiras da área de Computação, formando um corpus de 58 artigos em português, objetivando um levantamento de padrões morfossintáticos (combinações de categorias gramaticais) das palavras-chave elaboradas pelos autores dos artigos. Ressalta-se que não foi feito qualquer julgamento prévio sobre a adequação das escolhas dos autores. As revistas utilizadas foram:

- Revista Brasileira de Informática na Educação (4 exemplares)
- BC – Revista Brasileira de Computação - edições antigas, em português (8 exemplares)

Foram analisados 58 artigos, com suas respectivas palavras-chave definidas pelos autores. Os autores sugeriam, em média, quatro palavras-chave para um artigo, totalizando um conjunto de 219 palavras-chave. Os padrões mais freqüentemente encontrados foram:

- Nome;
- Nome Preposição Nome;

acompanhados por suas derivações adjetivadas:

- Nome Adjetivo;
- Nome Adjetivo Adjetivo;
- Nome Adjetivo Preposição Nome;
- Nome Preposição Nome Adjetivo;

onde 'Nome' trata-se de um Nome Próprio ou um Substantivo Comum.

Decidimos por utilizar esses padrões pelo fato dos mesmos cobrirem 189 das 219 palavras-chave analisadas, portanto, cerca de 86.3 %.

3.2 O Algoritmo EPC-P

Como o EPC-P necessita das classes gramaticais das palavras do texto para detectar os padrões, ele não trabalha sobre o texto original, mas sobre um texto etiquetado, onde todas as palavras aparecem associadas às suas categorias gramaticais. Dessa forma, o texto é pré-processado por um etiquetador (Part-of-Speech *Tagger*) de português produzido no NILC (Aires et al., 2000).

O texto etiquetado é percorrido em sua totalidade e são construídas seis listas, cada uma contendo todas as palavras do texto (em ordem alfabética) que se encaixam em um dos seis padrões procurados (nome, nome adjetivo, nome preposição nome, nome adjetivo preposição nome, nome preposição nome adjetivo, nome adjetivo adjetivo), assim como o número de vezes que cada uma das palavras ocorre no texto. É armazenado também o número de vezes que cada padrão se repete.

Paralelamente à construção dessas seis listas, são construídas mais seis listas, uma para cada padrão, sendo que nestas são inseridos somente os radicais das palavras. Estes radicais são obtidos utilizando-se o programa de extração de radicais (stemmer) de Porter (Porter, 1980), versão português, que identifica sufixos e prefixos, utilizando-se de regras heurísticas para isolar os radicais das palavras. Uma adaptação feita foi que desconsideramos os radicais das preposições, por serem de classe fechada, ou seja, não carregarem conteúdo semântico. Deste modo, por exemplo, "conceitos visuais em radiologia" e "conceitos visuais para radiologia" possuem os mesmos radicais, computando duas ocorrências de uma mesma frase, a despeito da diferença das preposições.

Feito o percurso do texto todo, ordenam-se as seis listas de radicais, deixando-as na forma decrescente do número de ocorrências de cada palavra no texto, enquanto que as seis listas que contêm as palavras originais continuam ordenadas alfabeticamente.

Terminada a fase de construção das doze listas, podemos nos desprender do texto e iniciarmos a construção da lista de palavras-chave. Primeiramente, criamos uma lista que damos o nome de "lista1". Preenchemos esta lista da seguinte forma: mantemos um ponteiro associado ao primeiro elemento de cada uma das seis listas de radicais; para cada um dos seis elementos analisados, dividimos o seu número de ocorrências pelo número de ocorrências do padrão em que ele se encontra (o que chamamos de freqüência relativa); o radical que tiver a maior freqüência relativa (e que ocorrer pelo menos duas vezes no texto) será inserido na *lista1*, juntamente com um marcador para indicar de qual lista ele foi retirado; repetindo o processo até que a *lista1* possua 50 elementos (número este escolhido por se mostrar suficiente para a obtenção das 30 palavras-chave finais, no próximo passo do algoritmo). Decidimos por utilizar a freqüência relativa ao invés de simplesmente escolher a palavra que ocorre mais vezes, porque concluímos que não se pode analisar simplesmente o número de vezes que a palavra ocorre sem levar em conta seu padrão, já que padrões como "nome" e "nome adjetivo" ocorrem bem mais vezes do que padrões como "nome adjetivo preposição nome" e "nome adjetivo adjetivo", que certamente não seriam considerados se o segundo método fosse utilizado. Por exemplo, este método considera o fato de "formação de recursos humanos" aparecer sete vezes no texto ser mais relevante do que "recursos humanos" aparecer dez vezes, o que nos parece mais razoável.

Depois de construída a *lista1*, iniciamos a construção da lista final de palavras-chave, a qual damos o nome de "lista2". A *lista2* é preenchida da seguinte maneira: pegamos o primeiro elemento da lista de radicais de nomes; pegamos a primeira ocorrência deste radical na *lista1*, se é que ele ocorre na mesma (desta forma, consideramos as primeiras palavras da lista de nomes como sendo as mais relevantes para o texto, e encontramos a melhor ocorrência de cada uma delas considerando todos os padrões); inserimos o radical encontrado na *lista2*, junto com o marcador que indica a classe a qual este elemento pertencia; pegamos o próximo elemento da lista de radicais de nomes; repetimos o processo (inserindo um novo radical na *lista2* sempre que ele ainda não pertencer à mesma) até que a *lista2* tenha 30 elementos, ou que acabe a lista de radicais de nome, o que faz com que a lista de palavras-chave não tenha necessariamente trinta elementos (Considerando que os autores costumam gerar de 3 a 6 palavras-chave para seus artigos, decidimos inicialmente por gerar 15 palavras-chave para cada artigo analisado. Porém, para que seja feita uma análise mais ampla da performance do

algoritmo, dado que ele se encontra em sua etapa inicial de construção, ampliamos este número para 30).

A *lista1* poderia ser considerada a lista final de palavras-chave, porém ela possui algumas imperfeições. A principal delas é que acabam por coexistir nela coisas muito parecidas, que poderiam ser resumidas a apenas uma delas. Por exemplo, obtivemos na *lista1* em um dos testes as frases "projeto cooperativo na Internet", "participação em projetos cooperativos", "projeto cooperativo". Com o método utilizado para a criação da *lista2*, "projeto cooperativo na Internet" foi considerada a melhor forma em que aparece o radical "projet" e, portanto, a única relevante entre as três. Esta "limpeza" na lista acaba por liberar espaço para que outros temas que também forem importantes sejam citados na lista. No caso citado anteriormente, por exemplo, "participação em projetos cooperativos" só seria considerado relevante pelo método utilizado se a palavra "participação" também tiver uma certa importância no texto.

Só nos resta então recuperar as palavras originais da *lista2*, já que ela é composta apenas por radicais. Isto é simples, já que, para cada elemento da *lista2*, temos armazenado um marcador que indica o padrão em que ele se encaixa, ou seja, a lista à qual ele pertencia. Basta então, para cada elemento, percorrer a lista de palavras originais correspondente ao padrão em que ele se encaixa, recuperando a melhor ocorrência deste radical na lista. Com isto, temos uma lista com até trinta palavras-chave para o texto.

4 EPC-R - EXTRAÇÃO DE PALAVRAS-CHAVE BASEADA EM FREQÜÊNCIA DE RADICAIS

Neste segundo método, que simplesmente chamamos de EPC-R (Extrator de Palavras-Chave por freqüência de Radicais), procuramos nos basear no algoritmo Extractor (Turney, 1999), para o inglês, que utiliza somente a freqüência de radicais no texto, não se prendendo a uma análise morfossintática das palavras do texto, diferentemente do algoritmo EPC-P.

4.1 Extractor

Extractor é um algoritmo de extração de palavras-chave por medição da freqüência de radicais no texto. Ele conta quantas vezes seqüências de palavras (simples, duplas ou trios) ocorrem no texto, utilizando-se de métodos estatísticos e de listas de stop-words (palavras que podem ser ignoradas no processo; em geral aquelas de classe fechada – conjunções, artigos, preposições - pontuação, etc.), para decidir quais delas são mais relevantes para o texto.

Seus bons resultados para o inglês nos levaram a utilizá-lo como parâmetro para construir um extrator de palavras-chave por freqüência de radicais para o português (Turney, 1997).

Extractor determina alguns dos parâmetros que usa por meio de algoritmos de aprendizado de máquina. Como não era nossa intenção perseguir o mesmo caminho, decidimos utilizá-los tal como aparecem, mesmo sabendo que tais valores carregam consigo uma dependência do corpus a partir do qual foram determinados. O objetivo é, posteriormente, investigar melhor o quanto sua variação influencia nos resultados produzidos pelos algoritmos para o português.

Por outro lado, parâmetros como FIRST_LOW_THRESH (representa até quando - que posição do texto - uma palavra deve ocorrer pela primeira vez para ser considerada relevante) e FIRST_HIGH_THRESH (representa depois de quando uma palavra deve ocorrer pela primeira vez para ser considerada uma ocorrência tardia) foram alterados, para se acomodarem melhor aos padrões de tamanho dos textos em português. Para obtenção dos mesmos, foi realizada uma pesquisa no corpus de artigos científicos, calculando-se o tamanho médio das Introduções dos artigos (por considerar que se uma dada palavra é relevante para o texto, ela

deve aparecer na introdução[1]), e chegamos aos valores: 450 e 800, respectivamente. Isso significa, portanto, que palavras relevantes tendem a ocorrer pela primeira vez até a posição 450, e que palavras que apenas ocorrem pela primeira vez após a posição 800 podem ser consideradas irrelevantes. Percebe-se a dependência desses valores ao corpus utilizado.

Outra diferença entre o Extractor e o EPC-R decorre da utilização das listas de stop-words. No EPC-R, decidimos utilizar duas listas de stop-words, para que uma consideração especial fosse feita quanto às preposições (explicada na seção seguinte).

Adicionalmente, o Extractor faz o uso de uma lista de verbos mais freqüentes do inglês (ele não trabalha com etiquetas morfossintáticas, de modo que precisa reconhecer essa categoria de outra forma). Decidiu-se não proceder da mesma forma para o português, uma vez que um algoritmo de reconhecimento morfológico das formas verbais não seria simples e nos desviaria de nossos objetivos. Essa diferença, no entanto, teve um impacto sensível no desempenho do EPC-R.

4.2 O Algoritmo do EPC-R

O EPC-R não precisa das classes gramaticais das palavras do texto, apenas torna-se necessário fazer algumas modificações no texto original, separando as pontuações das palavras do texto, para que cada palavra seja identificada com uma maior facilidade.

Tendo-se em mãos o texto adaptado, este é percorrido em sua totalidade e são construídas três listas, a primeira contendo todas as palavras simples do texto, a segunda, todas as duplas, e a terceira, todos os trios de palavras do texto (em ordem alfabética). As listas armazenam também o número de vezes que cada uma das palavras ocorre no texto.

Como neste método nós não temos conhecimento algum sobre a classe gramatical de cada palavra do texto, torna-se necessário utilizar uma lista contendo as stop-words mais freqüentes, para que, conforme o texto seja percorrido, torne-se possível que palavras sem importância (como artigos, conjunções, etc.) sejam descartadas.

Na verdade, utilizamos duas listas de stop-words, uma delas contendo as preposições junto com as demais palavras sem conteúdo semântico, e a outra sem as preposições. Utilizamos essas duas listas para que os trios de palavras possam possuir preposições como a palavra do meio. A idéia é que, por exemplo, "gerenciamento de software" seja uma candidata válida, mas que o mesmo não ocorra com "de software educacional" (já que, neste caso, a preposição "de" ocorre no começo do trio). Ou seja, neste algoritmo sempre descartamos as preposições, a não ser que ocorram como ligações entre duas palavras.

Paralelamente à construção destas três listas, são construídas mais três listas (correspondentes às listas de simples, duplas e trios), sendo que nestas são inseridos somente os radicais das palavras. Estes radicais são igualmente obtidos pelo *stemmer* de Porter (Porter, 1980).

Faremos agora uma análise referente às listas de radicais. Para cada radical, se ele aparecer pela primeira vez antes de FIRST_LOW_THRESH, multiplica-se sua freqüência por um valor que a valoriza (FIRST_LOW_FACTOR, 2.0). Se o radical aparecer depois de FIRST_HIGH_THRESH, então se multiplica sua freqüência por um valor que a desvalorize (FIRST_HIGH_FACTOR, 0.65). Um outro ajuste que precisamos fazer refere-se ao número de radicais nas candidatas a palavras-chave. Se a candidata tiver apenas um radical, nada é feito. Se ela tiver dois radicais, valoriza-se a mesma multiplicando-se sua freqüência por FACTOR_TWO_ONE (2.33). Se a palavra tiver três radicais, multiplicamos sua freqüência por FACTOR_THREE_ONE (5.00), valorizando-a ainda mais. O efeito dessa valorização é a preferência por frases mais completas sobre as mais simples.

---

[1] Uma Introdução de artigo científico deve fornecer o cenário, a relevância e os objetivos do trabalho, apontar a metodologia utilizada e adiantar sobre a natureza dos resultados obtidos.

Feito o percurso do texto todo, ordenamos cada uma das três listas de radicais, deixando-as na forma decrescente de freqüência, enquanto que as três listas que contêm as palavras originais continuam ordenadas alfabeticamente.

Terminada a fase de construção das seis listas, passamos à construção da lista de palavras-chave. Primeiramente, criamos a "lista1" com os 50 elementos de maior freqüência das três listas de radicais. A *lista2* é então preenchida da seguinte maneira: pegamos o primeiro elemento da lista de radicais simples; pegamos a primeira ocorrência deste radical na *lista1*, se é que ele ocorre na mesma (desta forma, consideramos as primeiras palavras da lista de radicais simples como sendo as mais relevantes para o texto, e encontramos a melhor ocorrência de cada uma delas considerando todas as três listas); inserimos o radical encontrado na *lista2*; pegamos o próximo elemento da lista de radicais simples; repetimos o processo (inserindo um novo radical na *lista2* sempre que ele ainda não pertencer a mesma) até que a *lista2* tenha trinta elementos, ou que acabe a lista de radicais simples (o que faz com que a lista de palavras-chave não tenha necessariamente trinta elementos).

Como no caso do EPC-P, a *lista1* poderia ser considerada a lista final de palavras-chave, mas nela ainda podem coexistir frases redundantes.

Resta então recuperar as palavras originais da *lista2*, já que ela é composta apenas por radicais. Isto é simples, já que para cada elemento da *lista2* basta contar por quantos radicais ele é composto para saber a que lista ele pertencia. Basta então, para cada elemento, percorrer a lista de palavras originais correspondente ao padrão em que ele se encaixa, recuperando a melhor ocorrência deste radical na lista. Com isto, temos uma lista com até trinta palavras-chave para o texto.

5 COMPARAÇÃO ENTRE EPC-P e EPC-R

Nesta seção, mostramos alguns dos resultados obtidos com os dois algoritmos.

Na Tabela 1, colocamos as palavras-chave obtidas pelos algoritmos, em comparação às elaboradas pelos autores de três artigos (Os resultados dos demais quinze artigos testados podem ser encontrados no Apêndice I).

Como nos testes realizados fixamos o número de palavras-chave a serem obtidas pelo programa em trinta (número este muito grande em comparação ao número de palavras-chave elaboradas pelos autores de cada artigo), torna-se irrelevante analisarmos o *precision* (porcentagem das palavras-chave geradas pelo algoritmo que também são consideradas chave pelo autor). Tendo isto em vista, fizemos apenas uma análise do recall (porcentagem das palavras-chave do autor que foram encontradas pelo algoritmo).

Tabela 1 – Resultados de Três Textos

Texto 1 **Título:** Agentes de Software para Análise das Interações em um Ambiente de Ensino à Distância. **Autores:** Patrícia Augustin Jaques, Flávio Moreira de Oliveira. **Palavras-chave do autor:** Educação à Distância, Sistemas Multi-Agentes, Ensino Colaborativo à Distância, Internet. **Palavras-chave encontradas pelo EPC-P:** agentes No sistema multi-agentes, mensagem de e-mail, Aluno, ambiente telemático de ensino, sistema Multi-agentes proposto, Lista de discussão, agente do professor, Análise das Interações, ASSUNTOS, informação, tipo de associação, ferramentas de comunicação, banco de dados, Educação à Distância, tarefa do professor, Trabalhos Futuros, Católica do Rio, Ensino à Distância, separação física, reunião de chat, conceitos de Educação, conjunto de agentes, endereço do arquivo, palavras-chave própria, participação dos alunos, Arquitetura geral do Sistema, Grande do Sul, Sala de Aula, salto de qualidade, caso do chat. **Recall EPC-P:** 0.25 **Palavras-chave encontradas pelo EPC-R:** Agent, Aluno, professor, Ambiente de Ensino, COMUNICAÇÃO, Educação à Distância, Sistema Multi-Agentes, Análise das Interações, conhecem, Ensino à Distância, 97, deve, outra, pode, ASSUNTOS, irá, Todas, dados, Lista de discussão, Colaboração, informação, Internet, Grande do Sul, mensagem de e-mail, Sistema Multi-Agentes proposto, Ambientes Telemáticos, encontram em andamento, ferramentas de comunicação, dados a partir, Sala de Aula.

Recall EPC-R: 0.75
Texto 2 **Título:** Design centrado no usuário ou no aprendiz? Uma proposta prática. **Autores:** Marcos Augusto Francisco Borges, M. Cecília C. Baranauskas2. **Palavras-chave do autor:** Design Centrado no Usuário, Design Centrado no Aprendiz, Aumento de Inteligência. **Palavras-chave encontradas pelo EPC-P:** Sistemas, design centrado no usuário, processo de design, design centrado no aprendiz, necessidades dos usuários, Interfaces, Jonas, ambiente Enxuto, Computação, proposta inicial, objetivos educacionais, desenvolvimento, sistemas baseados, mecanismo de scaffolding, melhoria do modelo, abordagem, adaptação das técnicas, construção de sistemas, dupla de funcionários, realização de experimentos, elemento central, funcionamento do sistema, realização da tarefa, treinamento da fábrica, Aumento de Inteligência, sistema especialista, folha de papel, poder computacional, sistema Enxuto, interface do software. **Recall EPC-P:** 1.00 **Palavras-chave encontradas pelo EPC-R:** Sistemas, usuário, processo de design, Aprendizado, Interfaces, não, pode, Computação, base, desenvolve, proposta, DCU, objetivo, centrado no usuário, necessidades dos usuários, utiliza, funcionários, conhece, Segundo, Técnicas, abordagem, computacional, 990, Soloway et al., não apenas, Desta, enquanto, Aprender, construída, contexto. **Recall EPC-R:** 0.00
Texto 3 **Título:** Um Framework para Estudo de Ambientes de Suporte à Aprendizagem Cooperativa. **Palavras-chave do autor:** CSCL, Ambientes de Aprendizagem Cooperativa Apoiados por Computador. Palavras-chave encontradas pelo EPC-P: ambientes de aprendizagem cooperativa, Aprendizagem Cooperativa, Teorias de Aprendizagem, Tipos, Desenvolvimento, Interação social, Atividades, tarefas cooperativas, construção de conhecimento, processos mentais, estudantes, aquisição de conhecimento, Domínios, forma, membros do grupo, objetivo, perspectiva, Aspectos, Computador, Modelos de Cooperação, trabalho, Áreas, desenvolvimento cognitivo, Cognição, comportamento, comunidade, contexto, elaboração, Framework, indivíduos. **Recall EPC-P:** 0.00 **Palavras-chave encontradas pelo EPC-R:** Aprendizagem, Ambiente, Aprendizagem Cooperativa, conhecer, estudantes, Outras, Atividades, pode, Desenvolvimento, grupo, Interação, objetivo, Não, forma, Tarefa, processo, Cognitiva, Dados, Trabalho, Memória de grupo, relacionada, Aprendiz, Desta, Framework, informação, Perspectivas, sobre, Estudo, diferente, problema. **Recall EPC-R:** 0.00

Mesmo assim, pensamos que uma análise mais subjetiva dos resultados deve ser feita. Por exemplo, no terceiro texto, apesar do recall atingido por ambos os métodos ser zero, vemos que isto ocorreu graças à má escolha das palavras-chave por parte dos autores ("CSCL" aparece apenas três vezes no texto e "Ambientes de Aprendizagem Cooperativa Apoiados por Computador" trata-se de uma frase muito grande para ser encontrada com exatidão – não se encaixa nos padrões do EPC-P e nem pode ser encontrada pelo EPC-R, por ser uma seqüência de mais de três palavras) e que os resultados obtidos pelo algoritmo, apesar de não coincidirem com as palavras-chave do autor, foram bem representativos.

No segundo texto, vemos uma limitação do EPC-R. No caso, ele não consegue encontrar as palavras-chave do autor por estas possuírem mais de três palavras ("design centrado no usuário", "design centrado no aprendiz"), apesar de chegar perto disso (encontrou "centrado no usuário", "processo de design", "Aprender"). Neste exemplo, vemos que o EPC-P teve um recall de 100%.

Analisando os resultados do primeiro texto, vemos que o EPC-R teve um recall melhor, ou seja, encontrou mais palavras-chave do autor do que o EPC-P, porém, o EPC-R acabou por encontrar também muitas palavras sem importância, como "97", "deve", "outra", "pode", entre outras. O EPC-P, apesar de ter um recall pior, manteve suas palavras-chave mais próximas ao tema central do texto e às palavras-chave do autor. Atribuímos isto ao fato de o EPC-P estar preso a padrões morfossintáticos, evitando, por exemplo, considerar relevantes verbos e numerais. Consideramos, portanto, que por um lado isto é bom para o EPC-P, já que faz com que quase toda sua lista de palavras-chave seja composta por palavras bem estruturadas (ao contrário do EPC-R), porém, por outro lado é ruim, já que faz com que o método não encontre algumas palavras importantes por estas não se encaixarem nos padrões.

Consideramos, assim, que para que seja possível a escolha do método de extração de palavras-chave a ser utilizado, o objetivo final da aplicação também deve ser levado em

consideração. Por exemplo, é mais importante utilizar um método de extração que atinja alto recall ou um que gere palavras-chave mais representativas, mesmo sendo distintas das do autor? (vale notar que em muitas aplicações essa informação nem estará disponível). No cenário de sumarização extrativa, por exemplo, consideramos o EPC-P mais eficiente, uma vez que quase a totalidade das palavras geradas possui um bom valor representativo, segundo uma análise absoluta, ou seja, não considerando o texto original, mas o grau de informação que ela carrega consigo.

Nas Tabelas 2 e 3, comparamos os recall e precision atingidos pelos dois métodos em dezoito textos. São mostradas também quantas palavras-chave o autor escolheu para cada texto (coluna N), quantas destas foram encontradas com exatidão pelos algoritmos (coluna NE), assim como as posições das palavras-chave do autor que foram encontradas na lista ordenada que é gerada por cada algoritmo (Pos). Assim, no Texto 7, por exemplo, duas das cinco palavras-chave do autor foram geradas pelo EPC-P (aparecendo nas posições 8 e 11 entre as 30 palavras-chave geradas). Portanto, uma precisão de 40% e uma recuperação de 6.7%.

Tabela 2 – Recall e Precision do EPC-P para 18 textos

	N	NE	Pos	Recall	Precision
Texto1	4	1	14	¼	1/30
Texto2	3	3	2, 4, 25	3/3	3/30
Texto3	3	1	4	1/3	1/27
Texto4	3	1	4	1/3	1/30
Texto5	4	0	-	0	0
Texto6	5	1	3	1/5	1/30
Texto7	5	2	8, 11	2/5	2/30
Texto8	3	1	4	1/3	1/27
Texto9	4	2	1, 24	2/4	2/30
Texto10	3	0	-	0	0
Texto11	6	1	9	1/6	1/30
Texto12	7	0	-	0	0
Texto13	4	2	1, 3	2/4	2/27
Texto14	5	0	-	0	0
Texto15	4	0	-	0	0
Texto16	3	1	6	1/3	1/27
Texto17	4	0	-	0	0
Texto18	7	1	25	1/7	1/30
TOTAL	**77**	**17**	*	**22.08%**	*

Tabela3 – Recall e Precision do EPC-R para 18 textos

	N	NE	Pos	Recall	Precision
Texto1	4	3	6, 7, 22	3/4	3/30
Texto2	3	0	-	0	0
Texto3	3	1	1	1/3	1/30
Texto4	3	1	5	1/3	1/30
Texto5	4	0	-	0	0
Texto6	5	2	7, 12	2/5	2/30
Texto7	5	3	12, 28, 29	3/5	3/30
Texto8	3	1	1	1/3	1/30
Texto9	4	1	4	1/4	1/30
Texto10	3	1	6	1/3	1/30
Texto11	6	1	5	1/6	1/30
Texto12	7	1	6	1/7	1/30
Texto13	4	1	2	1/4	1/30
Texto14	5	0	-	0	0
Texto15	4	1	3	1/4	1/30
Texto16	3	1	24	1/3	1/30
Texto17	4	1	1	1/4	1/30
Texto18	7	0	-	0	0
TOTAL	**77**	**19**	*	**24.68%**	*

Analisando as duas tabelas acima, vemos que o EPC-P encontrou com exatidão 17 das 77 (22.08 %) palavras-chave do autor, enquanto o EPC-R encontrou 19 (24.68 %). Definimos encontrar a palavra-chave do autor com exatidão como sendo quando o radical da palavra encontrada coincide com o radical da palavra dada como chave pelo autor (radical este encontrado pelo stemmer de Porter). Esta limitação torna esse tipo de análise do desempenho do algoritmo menos eficiente do que uma análise mais subjetiva, pois não faz nenhuma consideração quanto a sinônimos ou expressões de significado semelhante.

Voltando à análise das tabelas 2 e 3, no caso do EPC-P, verificamos que 70.58 % das palavras-chave geradas que correspondiam a alguma palavra-chave do autor (12 das 17) estavam entre as 10 consideradas mais relevantes pelo algoritmo, enquanto que apenas 17.64% (3 das 17) estavam entre as 15 consideradas menos relevantes.

Já fazendo uma análise dos resultados do EPC-R, verificamos que 68.42 % das palavras-chave geradas que correspondiam a alguma palavra-chave do autor (13 das 19) estavam entre as 10 consideradas mais relevantes pelo algoritmo, enquanto que 21.05% (4 das 19) estavam entre as 15 consideradas menos relevantes.

Pelo fato das quinze palavras iniciais geradas pelos programas parecerem mais relevantes, pensamos que quinze seria um valor razoável para o número de palavras-chave geradas pelos algoritmos para cada texto.

6 CONCLUSÕES E TRABALHOS FUTUROS

Neste documento foram discutidas a implementação e análise de dois algoritmos de extração de palavras-chave de textos em português, utilizando técnicas extrativas simples.

O primeiro deles (EPC-P) faz uma análise da freqüência de determinados padrões morfossintáticos no texto, para decidir quais palavras podem ser utilizadas para representar o tema central do mesmo. Um dos problemas encontrados na sua utilização foi que, por estar preso aos padrões, acaba por não encontrar outras construções interessantes, mas fora dos padrões (sendo, por exemplo, escolhidas como palavras-chave pelo autor). Um lado positivo de sua utilização foi que, também por estar preso aos padrões, acaba desconsiderando coisas que certamente não têm importância alguma, como verbos, numerais e pronomes.

Uma perspectiva para a sua melhoria seria a utilização de mais padrões, ou de recursos para verificação de sinônimos, como um thesaurus.

O segundo deles (EPC-R) faz uma análise da freqüência de radicais (simples, duplas ou trios) no texto, em detrimento à utilização de padrões. Um dos problemas encontrados em sua utilização foi o fato de ele estar preso à busca de, no máximo, trio de radicais. Outro provém do fato de ele não estar preso aos padrões, acabando por considerar relevantes algumas coisas que deveriam ser descartas, tendo o foco principal deste problema nos verbos (já que a tarefa de identificação de verbos no português é um tanto difícil). Porém, assim como no outro algoritmo, o fato dele não estar preso a padrões tem o seu lado positivo, pois isso fornece a ele uma variedade maior de padrão das palavras a serem geradas.

Uma perspectiva para a sua melhoria seria a possibilidade de busca de uma maior seqüência de radicais (quadras, quinas...), assim como a utilização de outras técnicas adicionais como a verificação de verbos, numerais, pronomes, entre outros, além da verificação de sinônimos.

Visto que o objetivo deste trabalho era dar os primeiros passos para a criação de um método de extração de palavras-chave para o português mais eficiente, consideramos os resultados obtidos bem satisfatórios.

AGRADECIMENTOS

Os autores agradecem o apoio financeiro do Programa PIBIC-CNPq (USP) e a colaboração dos pesquisadores do NILC.

REFERÊNCIAS BIBLIOGRÁFICAS

AIRES, Rachel V.X.; ALUÍSIO, S. M.; KUHN, D. C. S.; ANDREETA, M. L. B.; OLIVEIRA Jr., O. N. Combining Multiple Classifiers to Improve Part of Speech Tagging: A Case Study for Brazilian Portuguese. *(SBIA'2000)* Atibaia, SP, November, 20-22.

BLACK, W.J. and JOHNSON, F.C. Expert Systems for Information Management: a Practical Evaluation of Two-Based Automatic Abstraction Techniques. Department of Computaion. University of Manchester Institute of Science and Technology 1(3), 1998.

PORTER, M. F. *An algorithm for suffix stripping*. Program, 14(3):130--137, 1980.

TURNEY, Peter. *Learning to Extract Keyphrases from Text*, Tech. Report Number NRC-41622, National Research Council Canada, Institute for Information Technology, 1999.

TURNEY, Peter. *Extraction of keyphrases from text:Evaluation of four algorithms*. Technical report, Institute for Information Technology, 1997.

SOUZA, Carolina. *Algoritmos de Sumarização Extrativa de Textos em Português*. Relatórios Técnicos do ICMC-USP. Setembro, 2001.

SPARCK JONES, Karen. *Discourse Modelling for Automatic Summarising*. Tecnhical Report 290, University of Cambrige, 1993.

APÊNDICE I

TESTES

Texto 4:
Título: Estratégias de Ensino em Sistemas Tutores Inteligentes modelados através da tecnologia de agentes
Autores: Lucia Maria Martins Giraffa, Rosa Maria Viccari.
Palavras-chave do autor: *Sistemas Tutores Inteligentes, Estratégias de Ensino, Agentes pedagógicos.*
Palavras-chave encontradas pelo EPC-P: aluno, Tutor, Sistemas, Estratégias de Ensino, processo mentais, trabalho do aluno, Artificial, forma, Proceedings, Agents, Intelligence, Conference, jogo educacional, ações do aluno, uso de recursos, agentes reativos, modelo de estudante, Exemplos de regras, projeto de STI, nível de energia, ensino utilizadas, conjunto de estados mentais, conhecimento diferenciados, crença do tutor, comportamento do tutor, elemento do cenário, grau de interferência, pesquisadores da área, tempo suficiente, implementação de STI.
Recall EPC-P: 0.33
Palavras-chave encontradas pelo EPC-R: aluno, Tutor, utiliza, Sistemas, Estratégias de Ensino, trabalham, Agent, Ambiente, não, on Artificial Intelligence, estudante, Model, através, onde, pode, processo, Projeto, pedagógica, simulam, STI, Artificial Intelligence in, mentais, Exemplo, faz, combinação de técnicas, pesquisadores da área, aplicada à Educação, Proceedings... Kobe, Pedagogical Agents, Conference on Artificial.
Recall EPC-R: 0.33

Texto 5:
Título: Modelagem do Aprendiz baseado no Paradigma de Ensino Cooperativo
Autores: Jeane Silva Ferreira, Sofiane Labidi.
Palavras-chave do autor: *Cooperative Tutoring System,Student Modeling,Cooperative Learning Supported Computer, Multi-Agents System.*
Palavras-chave encontradas pelo EPC-P: Modelo Individual do Aluno, Modelo Conceitual do Sistema, grupo, banco cognitivo, Aprendizagem Cooperativa, componente do conhecimento específico, Modelagem, professor, trabalho, processo de ensino/aprendizagem, comportamento dos alunos, estratégias pedagógicas cooperativas, respostas corretas, Sistemas de Tutoria, módulo de avaliação, área cooperativa, características cognitivas do aluno, Ensino cooperativo, Modelo de Aprendiz, modelo do domínio, banco de erros, banco de hipóteses, modelo de aluno proposto, sessão de aprendizagem cooperativa, banco de acerto, modelo de aluno composto, funcionamento do modelo individual.
Recall EPC-P: 0
Palavras-chave encontradas pelo EPC-R: Aluno, Modelo, Cooperativa, Sistema, Aprendizagem Cooperativa, conhecer, professor, componente, estudante, não, pode, Sistemas de Tutoria, Modelagem do Aprendiz, Apresentamos, grupo, et al., resultado, Computador, estratégia, desenvolvido, Paradigma de Ensino, Inteligente, objetivo, outra, Todos, aula, comportamento, forma, processo de ensino/aprendizagem, baseada no paradigma.
Recall EPC-R: 0

Texto 6:
Título: Projeto KidFamília Br. A Democratização da Internet, A Socialização do Jovem Carente e a Recuperação da Auto-Estima do Idoso.
Autora: Marisa Lucena
Palavras-chave do autor: *Comunidades Dinâmicas para a Aprendizagem, Internet, KidFamília, Khouse, Kidlink.*
Palavras-chave encontradas pelo EPC-P: Projeto, Crianças, KHOUSE, Computador, Kidlink no Brasil, atividade, Rio de Janeiro, cultura familiar, listas de interesse, Br, Educação à Distância, interesse geral, Democratização da Internet, M., anos de idade, casa, centro, Alunos da Escola, LUCENA, Brasileira de Informática, pesquisas em temas, Características no Brasil, trabalho cooperativo, uso de computadores, segmentos da Sociedade, Construção do Conhecimento, José dos Campos, cultura do aprendizado, sub-projeto no Brasil, biblioteca pública.
Recall EPC-P: 0.20
Palavras-chave encontradas pelo EPC-R: Projeto, Kidlink no Brasil, Computador, Crianças, Pode, Não, Internet, Escola, Faça, Educação, Trabalhar, KHOUSE, ativa, Novo, apresenta, listas de interesse, Jovens, sobre, Br, Todos, Desenvolvida, Deste, Experiência, LUCENA, participam, Pesquisa, anos de idade, 5 anos, Algumas, cultura de aprendizado.
Recall EPC-R: 0.40

Texto 7:
Título: Suporte de Rede para as Aplicações de Realidade Virtual destinadas a Educação à Distância
Autores: Mario Massakuni Kubo, Felicio Vanderlei Deriggi Jr., Antônio Carlos Sementille, Claudio Kirner.
Palavras-chave do autor: *Realidade Virtual, Ambientes Virtuais Distribuídos, CORBA, Multicast, Educação à Distância.*
Palavras-chave encontradas pelo EPC-P: Objeto Servidor de E/S, cliente, mensagem, Serviço de Eventos, aplicação desenvolvida, Suporte de Rede, sala, CORBA, Arquitetura do Suporte, comunicação multicast, Educação à Distância, informação, participante, Gerenciamento do AV, J., Modelos de Comunicação, OrbixTalk, long, M., utilização da plataforma, gargalo do sistema, Objetos de Gerenciamento, trabalhos futuros, Ambientes Virtuais, construção do suporte,

saída de usuários, técnica multicasting, Canal de Eventos, controle de entrada, Ferramenta de Apoio.
Recall EPC-P: 0.40
Palavras-chave encontradas pelo EPC-R: Objeto, Aplicação, mensagem, Desta, Através, Serviço de Eventos, Modelos de Comunicação, pode, Plataforma CORBA, Suporte de Rede, Virtual, Ambientes Virtuais Distribuídos, informa, Sistema, Professor, 996, cliente, outra, sobre, desenvolver, forma, RV, Deriggi, implementada, Interface, manipulado, Projeto, Educação à Distância, Realidade Virtual, computação distribuída orientada.
Recall EPC-R: 0.60

Texto 8:
Título: Educação e Informática: a construção de um projeto na rede pública de ensino
Autores: Sonia Schechtman Sette, Márcia Angela Aguiar, José Sergio Antunes Sette.
Palavras-chave do autor: *políticas de informática na educação, formação de recursos humanos, informática na educação.*
Palavras-chave encontradas pelo EPC-P: Informática na rede municipal, rede municipal de ensino, Projeto, Informática na Educação, Programa, Secretaria de Educação, área, Curso de Especialização, desenvolvimento, professores, formação de profissionais qualificados, unidade escolar, Política, processo, capacitadores, Espaço, experiência, Federal de Pernambuco, rede municipal de Recife, Brasileiro de Informática, áreas de conhecimento, Nacional de Informática, Núcleos Profissionalizantes de Informática, recursos humanos, conhecimento da tecnologia, ambiente favorável, projeto pedagógico.
Recall EPC-P: 0.33
Palavras-chave encontradas pelo EPC-R: Informática na Educação, Rede, Projeto, Escola, Ensino, Programa, área, desenvolvendo, Rede Municipal, Secretaria de Educação, Formação, Programa de Capacitação, processam, Experiências, Nessa, pedagógicas, Política, profissionais, conhecimento, diversas, docente, não, Tecnologia, Universidade Federal, Aos, forma, implantados, pública de ensino, vem sendo desenvolvida, Educacional.
Recall EPC-R: 0.33

Texto 9:
Título: AS TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO E A EDUCAÇÃO
Autor: QUARTIERO, E. M.
Palavras-chave do autor: *tecnologia, aprendizagem, trabalho pedagógico, conhecimento.*
Palavras-chave encontradas pelo EPC-P: TECNOLOGIAS, sociedade do conhecimento, sistema, EDUCAÇÃO, processo, trabalho, Tecnologias da informação, Informática, possibilidade, Técnico do SENAC, Professora, Rio de Janeiro, São, complexidade, meios de comunicação, Paulo, recurso, sujeito, alunas, Ciências, conjunto, conteúdo, discussão, aprendizagem, atividade de diálogo, conceito, construção do conhecimento, Engenheiro, forma, futuro.
Recall EPC-P: 0.50
Palavras-chave encontradas pelo EPC-R: conhecer, Pedagoga, EDUCAÇÃO, TECNOLOGIAS, trabalha, Informática, TECNOLOGIAS DA INFORMAÇÃO, COMUNICAÇÃO, Não, interativa, Novas, sobre, utilizadas, processo de interação, sistema, Computador, discussão, ensino, outra, possibilidade, escola, permite, cognitiva, dessa, forma, grande, meios de comunicação, recurso, Técnico, Todos.
Recall EPC-R: 0.25

Texto 10:
Título: Ambiente Hipertexto para Apoio ao Planejamento e Elaboração de Cursos de Treinamento Segundo a Metodologia TRAINAIR
Autores: José Maria Parente de Oliveira, Clovis Torres Fernandes.
Palavras-chave do autor: Apoio ao Planejamento de Cursos, Ambiente Hipertexto, Metodologia TRAINAIR.
Palavras-chave encontradas pelo EPC-P: tipo de entidade, fases prescritas na Metodologia, recuperação das informações compiladas, Projeto, desenvolvimento cooperativo de cursos, Instanciação, formulário, processador, ambiente, estrutura interna dos tipos, tipos de elos aplicativos, TRAINAIR, arquitetura básica do TrainTool, componentes de entidades diferentes, conjunto, Preparação, Trabalho, Figura, Objetivos Intermediários, Concepção do Plano, Acompanhamento de Projeto, Centros de Instrução, material didático final, problemas identificados, ambiente de apoio, Plano de Estudos, atualizações necessárias em tópicos, tópicos relacionados de fases.
Recall EPC-P: 0
Palavras-chave encontradas pelo EPC-R: Fase, preparação de curso, Informação, TrainTool, Projeto, Metodologia TRAINAIR, Ambiente, Aplicação, formulário, Objetivo, Trabalho, forma, pode, Tipos de Entidades, Implementação, processador de textos, Cursos de Treinamento, definida, ambiente de apoio, Conjunto, ativadas, desenvolvido, parte, dado, Dessa, realizadas, Segundo a Metodologia, 2, Hipertexto, Centros de Instrução.
Recall EPC-R: 0.33

Texto 11:
Título: Astral: Um Ambiente para Ensino de Estruturas de Dados através de Animações de Algoritmos
Autores: Islene C. Garcia, Pedro J. de Rezende, Felipe C. Calheiros
Palavras-chave do autor: Estruturas de Dados, Animação de Algoritmos, Visualização de Algoritmos, Laboratório de Ensino, Ambientes de Aprendizagem Baseados em Computador.
Palavras-chave encontradas pelo EPC-P: implementação das animações gráficas, Algoritmos, Ambientes, Figura, camada específica, uso, estudante possa, usuário, Vantagens da abordagem construtiva, Estruturas de Dados, Projetos,

Brown, b, elemento, processamento, Utilização, visualizador, conjunto minimal, disciplina de laboratório, detecção de erros, objetos gráficos, ambiente de programação, Ensino de Estruturas, gerenciador de eventos, Compreensão de Algoritmos, ferramenta de ensino, construção de animações, efeito colateral na remoção, interrupção da execução, estado da estrutura.
Recall EPC-P: 0.20
Palavras-chave encontradas pelo EPC-R: AnimA, Algoritmos, Ambiente, Implementadas, Estruturas de Dados, Podem, Utiliza, animação gráfica, estudante, Projetos, Abordagens, Astral, Brown, Deste, programa, Figura, uso, visualizador, seção,0, aprendiz, apresenta, Computação, desenvolvemos, Ensino de Estruturas, objetivo, sistema, 4, ativa, através.
Recall EPC-R: 0.20

Texto 12:
Título: COMUNIDADES DINÂMICAS PARA O APRENDIZADO NA INTERNET
Autora: Marisa Lucena
Palavras-chave do autor: Desenho Instrucional, Construtivismo, Internet, Comunidade Dinâmica para o Aprendizado, WWW, Listas de Interesse, Comportamento de Grupos
Palavras-chave encontradas pelo EPC-P: membro do grupo, Conhecimento, projeto cooperativo na Internet, Necessidade, Process, Informação, ambiente de aprendizagem, espaço ilimitado, forma, Lucena, M., objetivos educacionais, Rio de Janeiro, Resultados, APRENDIZADO, COMUNIDADES, Especialistas, nível, Atividades, pessoa, Compreensão do Comportamento, lista de discussão, Modelo de Escola, recurso tecnológico, aprendizagem dinâmica, participação em projetos cooperativos, Solução do Problema, sala de aula, Membros individuais, Kidlink no Brasil.
Recall EPC-P: 0
Palavras-chave encontradas pelo EPC-R: Conhecimento, Aprendizagem, Novo, Informação, Grupo, deve, INTERNET, Cooperativa, Process, Todas, COMUNIDADES DINÂMICAS, Lucena, Desenvolvimento, M., pode, APRENDIZADO, espaço, Necessidade, Objetivos, Outros, Neste caso, forma, Compartilhamento, &, vez, Controle, Projeto, determinada, membro do grupo, tecnologia.
Recall EPC-R: 0.14

Texto 13:
Título: Educação de Pacientes através de Sistemas de Acesso Público
Autores: Carla Valle, Antonio Augusto Ximenes, Gilda Helena B. de Campos, Ana Regina Rocha, Alvaro Rabelo Jr.
Palavras-chave do autor: educação de pacientes, sistemas de acesso público, multimídia, qualidade.
Palavras-chave encontradas pelo EPC-P: Sistemas de acesso público, Informação, Educação de Pacientes, característica do sistema, AVALIADOS, acesso público na educação, exame de cintilografia miocárdica, usuários finais, conjunto de atributos referentes, utilizabilidade do sistema, uso, tela sensível ao toque, atributos relacionados à facilidade, ponto de vista, Rio de Janeiro, caso de Sistemas, cardiologista responsável pelo Setor, produto de software, realização de exames, Clareza da Informação, confiabilidade conceitual, Bahiana de Cardiologia, Unidade de Cardiologia, alimentos proibidos, Setor de Medicina, atributos identificados, conjunto de atributos relacionados.
Recall EPC-P: 0.50
Palavras-chave encontradas pelo EPC-R: Sistema, Educação de Pacientes, Informação, Utilizabilidade,995, Sistema de Acesso, Medicações, Acesso Público, sobre, uso, Devo, Não, pode, relacionadas, &, aos, Bahiana de Cardiologia, fornece, et al., Outra, através, Computadores, considerado, problema, Usuários, casas, Exame, Desta,994, característica do sistema.
Recall EPC-R: 0.25

Texto 14:

Título: ESTADO DA ARTE EM ESPAÇOS VIRTUAIS DE ENSINO E APRENDIZAGEM
Autores: Neide Santos
Palavras-chave do autor: Tecnologias de rede e educação. Educação baseada na Web. Aprendizagem cooperativa apoiada por computador. Frameworks para aprendizagem cooperativa. Análise das formas de EBW.
Palavras-chave encontradas pelo EPC-P: Curso, Ambientes, Sistema, Site, Ferramenta, forma, Apoio, Professor, Estudante, Web, objetivo, Alunos, Espaços, Salas de aula virtuais, trabalho cooperativo, aprendizagem cooperativa, Internet, Análise, Suporte, apresentação da atividade educacional, usuário, grupo, resolução cooperativa de problemas, conteúdos curriculares, desenvolvimento de atividades cooperativas, educação virtual, Sistemas de Autoria, base de dados, recursos didáticos variados, espaço de comunicação.
Recall EPC-P: 0
Palavras-chave encontradas pelo EPC-R: Curso, Ambiente, Cooperativa, Podem, Sistemas de Autoria, ¨, Professor, Site, Desenvolvemos, Aprendizagem, Formas, Aluno, Analisamos, Não, Trabalho, Web, Apoiada, Internet, Virtual, apresenta, Acessamos, Informações, objetivo, comunicação, Educacionais, fornece, Suporte, página Web, permite, Sala de Aula.
Recall EPC-R: 0

Texto 15:
Título: Infra-estrutura de Suporte à Editoração de Material Didático Utilizando Multimídia
Autores: Maria Alice Soares de Castro, Rudinei Goularte, Elderclei Regis Reami, Edson dos Santos Moreira
Palavras-chave do autor: Multimídia, Hipermídia, Informática na Educação, Video-On-Demand
Palavras-chave encontradas pelo EPC-P: Sistema, servidor de dados multimídia, Professor, Usuário, Informação, Vídeo, aplicativos, cliente, Utilização de Multimídia, recursos, acesso, Aluno, apresentação de material didático, módulo,

processamento, componente, pedido, conjunto de hiperdocumentos, ferramenta, Hipertexto, funções, proposta de ambiente, pesquisa, Suporte, conteúdo semântico dos dados, mídia contínua, tempo real, maneiras diferentes, número de usuários, ligações conceituais.
Recall EPC-P: 0
Palavras-chave encontradas pelo EPC-R: Sistema, Usuário, Multimídia, pode, Informação, apresenta, Professor, Utilizando, Através, Devido, Apoio ao Aprendizado, não, Interação, Ambiente, recursos, Baseado, comunicadores, conhecem, Material Didático, Dados, Cada, diversas, ferramenta, Oferecer, Formas, implementadas, Interland, nova, processamento, fornece.
Recall EPC-R: 0.25

Texto 16:
Título: LICENCIATURA EM INFORMÁTICA - Uma questão em aberto
Autores: Sonia Schechtman Sette, Márcia Angela Aguiar, José Sergio Antunes Sette
Palavras-chave do autor: Formação de Recursos Humanos, Licenciaturas, Currículos
Palavras-chave encontradas pelo EPC-P: Política Nacional de Formação, conhecimento, INFORMÁTICA, ensino básico, objetividade, Curriculares, Cursos, professor, disciplina, Licenciatura em Informática, área de conhecimento, melhoria da educação básica, informática na escola, conteúdo, uso de software educacional, sociedade contemporânea, Competência linguística argumentativa, conceitos básicos sobre computadores, Profissional de Educação, editor gráfico, estudo das linguagens, formação de profissionais, setores do empresariado nacional, recursos humanos, reforma educacional, disciplina específica, conhecimento humano.
Recall EPC-P: 0.33
Palavras-chave encontradas pelo EPC-R: INFORMÁTICA, conhecer, Educação, Formação, Escola, LICENCIATURA EM INFORMÁTICA, cursos de licenciatura, objetividade, Sociedade, área, básica, aos, Desta, diversas, Educacional, Humanas, Pedagógica, sobre a formação, Competência, Docente, Nacional, Formação do Profissional, Tal, Curriculares, Desse, Ensino, campo, idéia, melhor, modo.
Recall EPC-R: 0.33

Texto 17:
Título: Pesquisa e Desenvolvimento em Informática na Educação no Brasil – Parte I
Autores: Clovis Torres Fernandes, Neide Santos
Palavras-chave do autor: Informática na Educação, Grupos de Pesquisa em Informática na Educação no Brasil, Linhas de Pesquisa e Projetos em Informática na Educação
Palavras-chave encontradas pelo EPC-P: Educativa, Projetos de Pesquisa, Brasileira de Informática, conceitos visuais com Sistemas, ambiente de aprendizagem telemático, Grupos Brasileiros de Pesquisa, desenvolvimento cognitivo, grupos atuantes na área, Linhas de Pesquisa, Editorial da Revista, José dos Campos, cursos hipermídia, Processos cognitivos, ensino de conceitos visuais, formação de recursos humanos, Informações complementares, aprendizagem mediados por computador, rede telemática, Educação à Distância, Lista Inicial dos Perfis, software educacional, Representações de Espaços Bidimensionais, canal deficientes, Projetos Em Andamento, Medidas cognitivas, ordem alfabética, Sistemas multiagentes, Hipermídias adaptativas.
Recall EPC-P: 0
Palavras-chave encontradas pelo EPC-R: Informática na Educação , Pesquisa , Desenvolver , Grupo de Pesquisa , Revista Brasileira, Brasileira de Informática , 1998. , Projetos de Pesquisa , área de Informática , Computadores , Ensino , 1., José dos Campos , Universidade , Informações Complementares, Tecnologia , Página Web, Ambiente de Aprendizagem , Clovis Torres Fernandes , Process , Publicações , Instituto , BRASIL , Linhas de Pesquisa , Mestrado , Principais Linhas, Ciência da Computação , atuantes na área , Educação à Distância , Perfis dos Grupos .
Recall EPC-R: 0.33

Texto 18:
Título: SOFTWARE EDUCATIVO MULTIMIDIA ASPECTOS CRÍTICOS NO SEU CICLO DE VIDA
Autores: Alvaro H. GALVIS-PANQUEVA, D. Ed.
Palavras-chave do autor: Engenharia de sofware educativo, planejamento estratégico de informática em educação, o computador como meio, informação multimidia, interatividade, micro-mundos interativos, ambientes multimidia de aprendizagem.
Palavras-chave encontradas pelo EPC-P: ambientes multimidia interativos, processamento, criação de ambientes educativos, INFORMÁTICA, Controle, MEC, aprendiz, sentido, solução de problemas, qualidades únicas do computador, MICROMUNDOS, tipo heurístico, Santafé de Bogotá, usuário, valor, base, campo vital, contexto, Marcos tecnológicos, meios dispostos, eixo focal, ambiente da aprendizagem, apoio informático, tipos de MECs, engenharia de sofware educativo, ciclo de vida, enfoque estratégico, Universidade de Los, redes virtuais, inteligencia artificial.
Recall EPC-P: 0.17
Palavras-chave encontradas pelo EPC-R: EDUCATIVO, Não, INFORMÁTICA, AMBIENTES, faz, processamento, Outros, usuário, MULTIMIDIA, deve, base, Cada, Computação, NECESSIDADES, sentido, tipo, campo, desenvolve, maneira, pode, MARCO, solução de problemas, trabalha, apenas, ESE, grande, Universidade de Los, Professor, cada vez, aprendiz.
Recall EPC-R: 0