

Análise do uso de *feedback* de relevância no Sistema de Integração Lattes-Qualis (SILQ)

Carlos Bonetti¹

¹Bacharelando de Ciência da Computação
Departamento de Informática e Estatística
Centro Tecnológico
Universidade Federal de Santa Catarina

Orientação: Profa. Dra. Carina F. Dorneles

Trabalho de Conclusão de Curso, 2016

Sumário

Introdução

- Histórico e Justificativa
- Objetivos

Conceitos

- Information Retrieval e Data-matching

Desenvolvimento

- Alterações tecnológicas
- Uso de feedback de relevância

Conclusões

Sumário

Introdução

Histórico e Justificativa

Objetivos

Conceitos

Information Retrieval e Data-matching

Desenvolvimento

Alterações tecnológicas

Uso de feedback de relevância

Conclusões

Histórico e Justificativa

- ▶ AGUIAR, Felipe Nedel de; COSTA, Maria Eloísa. **SILQ - Sistema de Integração Lattes Qualis**. Trabalho de Conclusão de Curso. Florianópolis: Universidade Federal de Santa Catarina, Biblioteca Universitária, 2015.

Histórico e Justificativa

- ▶ AGUIAR, Felipe Nedel de; COSTA, Maria Eloísa. **SILQ - Sistema de Integração Lattes Qualis**. Trabalho de Conclusão de Curso. Florianópolis: Universidade Federal de Santa Catarina, Biblioteca Universitária, 2015.
- ▶ Qualificação automática de produções científicas através de busca por similaridade textual nos dados Qualis;

Histórico e Justificativa

- ▶ AGUIAR, Felipe Nedel de; COSTA, Maria Eloísa. **SILQ - Sistema de Integração Lattes Qualis**. Trabalho de Conclusão de Curso. Florianópolis: Universidade Federal de Santa Catarina, Biblioteca Universitária, 2015.
- ▶ Qualificação automática de produções científicas através de busca por similaridade textual nos dados Qualis;
- ▶ <http://silq.inf.ufsc.br>

Fotinhas da primeira versão do SILQ



Lattes

Dados Gerais	Linhas de pesquisa	Projetos	Áreas	Produção em C.T. & A.	Bancas	Eventos	Orientações	Citações
<div>  <div> <p>André Azevedo da Fonseca</p> <p>Coordenador do curso de Comunicação Social da Universidade de Uberaba. Pós-doutorando em Estudos Culturais, no Programa Avançado de Cultura Contemporânea - PACCUPRI, Doutor em História pela Universidade Estadual Paulista - Julio Mesquita Filho (Unesp). Especialista em História do Brasil pela PUC-MG. Graduado em Comunicação Social (Jornalismo) pela Universidade de Uberaba (Unube). Autor do livro-reportagem "Cotidianos culturais e outras histórias: a cidade sob novas olhares" (Unube, 2004) e de quatro capítulos de livros. Como estudante universitário conquistou 8 prêmios nacionais. Como professor orientador, conquistou 19 prêmios nacionais e internacionais. Em 2010 foi contemplado no Prêmio Rumos Pesquisa Aplicada do Itaú Cultural. Tem experiência na área de Livro-reportagem, Docência Superior e pesquisa acadêmica nas áreas de Comunicação, Educação e História. Os interesses de pesquisa, com ênfase na interdisciplinaridade, englobam Comunicação, Cultura e Hegemonia, Movimentos Sociais e Culturais Contemporâneos, Indústria criativa, Estudos Culturais, Cultura Política, Política Cultural, Direitos Culturais, Imaginários Sociais e Políticos, Mitologias Políticas, Redes sociais, Comunicação e Educação, Comunicação e Cidadania. (Texto informado pelo autor)</p> <p>Última atualização do currículo em 26/01/2012</p> <p>Endereço para acessar este CV: http://lattes.cnpq.br/5565508482791214</p> </div> <div>  <p>Certificado pelo autor em 26/01/12</p> </div> <div> <p>Rede de colaboração</p> <p>Rede de grupos de pesquisa</p> <p>SciELO - artigos em texto completo</p> </div> </div>								
<p>Dados pessoais</p> <p>Nome André Azevedo da Fonseca</p> <p>Nome em citações bibliográficas FONSECA, A. A.</p> <p>Sexo Masculino</p> <p>Endereço profissional Universidade de Uberaba, Câmara de Ensino de Graduação, Pós-Graduação, Pesquisa e Extensão, Curso de Comunicação Social Av. Nere Sabino, 1801 - Bloco L Universitário 36055-500 - Uberaba, MG - Brasil Telefone: (34) 33109952 URL da Homepage: http://www.comunicacao.unube.br</p>								
<p>Formação acadêmica/Titulação</p> <p>2011 Pós-Doutorado - Universidade Federal do Rio de Janeiro, UFRJ, Brasil Bolsista do(a) Itaú Cultural</p>								

Qualis

ISSN	Título	Área de Avaliação	Estrato
1041-4347	IEEE Transactions on Knowledge and Data Engineering (Print)	CIÊNCIA DA COMPUTAÇÃO	►A1
0018-9464	IEEE Transactions on Magnetics	CIÊNCIA DA COMPUTAÇÃO	►B4
0278-0062	IEEE Transactions on Medical Imaging (Print)	CIÊNCIA DA COMPUTAÇÃO	►A1
1536-1233	IEEE Transactions on Mobile Computing	CIÊNCIA DA COMPUTAÇÃO	►A2
1520-9210	IEEE Transactions on Multimedia	CIÊNCIA DA COMPUTAÇÃO	►A2
2162-237X	IEEE Transactions on Neural Networks and Learning Systems	CIÊNCIA DA COMPUTAÇÃO	►A1
0018-9499	IEEE Transactions on Nuclear Science	CIÊNCIA DA COMPUTAÇÃO	►B1
1045-9219	IEEE Transactions on Parallel and Distributed Systems (Print)	CIÊNCIA DA COMPUTAÇÃO	►A2
0885-8950	IEEE Transactions on Power Systems	CIÊNCIA DA COMPUTAÇÃO	►B2
0098-5589	IEEE Transactions on Software Engineering	CIÊNCIA DA COMPUTAÇÃO	►A1
1083-4427	IEEE Transactions on Systems, Man and Cybernetics. Part A, S	CIÊNCIA DA COMPUTAÇÃO	►A2
1094-6977	IEEE Transactions on Systems, Man and Cybernetics. Part C, Ap	CIÊNCIA DA COMPUTAÇÃO	►A2
0018-9545	IEEE Transactions on Vehicular Technology	CIÊNCIA DA COMPUTAÇÃO	►A1
1063-8210	IEEE Transactions on Very Large Scale Integration (VLSI) System	CIÊNCIA DA COMPUTAÇÃO	►A2
1077-2626	IEEE Transactions on Visualization and Computer Graphics	CIÊNCIA DA COMPUTAÇÃO	►A2
1536-1276	IEEE Transactions on Wireless Communications	CIÊNCIA DA COMPUTAÇÃO	►A1
1536-1284	IEEE Wireless Communications	CIÊNCIA DA COMPUTAÇÃO	►A1
2162-2337	IEEE Wireless Communications Letters	CIÊNCIA DA COMPUTAÇÃO	►B4
1932-4537	IEEE eTransactions on Network and Service Management	CIÊNCIA DA COMPUTAÇÃO	►B3
1932-8540	IEEE-RITA	CIÊNCIA DA COMPUTAÇÃO	►B5
1545-5963	IEEE/ACM Transactions on Computational Biology and Bioinform	CIÊNCIA DA COMPUTAÇÃO	►B1
0916-8532	IEICE Transactions on Information and Systems	CIÊNCIA DA COMPUTAÇÃO	►B1
1751-861X	IET Computers & Digital Techniques (Online)	CIÊNCIA DA COMPUTAÇÃO	►B1
1751-8601	IET Computers & Digital Techniques (Print)	CIÊNCIA DA COMPUTAÇÃO	►B1
1751-8806	IET Software (Print)	CIÊNCIA DA COMPUTAÇÃO	►B1
1091-9856	INFORMS Journal on Computing	CIÊNCIA DA COMPUTAÇÃO	►B1
1526-5528	INFORMS Journal on Computing (Online)	CIÊNCIA DA COMPUTAÇÃO	►B1

Motivação

Atualização tecnológica e da base de dados

Hipótese

É possível aumentar a taxa de acerto do SILQ utilizando *feedback* de usuários?

Objetivos

Objetivo geral

Analisar o impacto que o uso de feedback de relevância tem na precisão dos resultados de avaliações realizadas pelo SILQ, efetuado sobre uma nova arquitetura da ferramenta que inclui a criação de API de integração com outros sistemas e a atualização da base de dados conforme as novas classificações Qualis.

Objetivos específicos

1. Reestruturação da arquitetura e banco de dados do SILQ a fim de suportar classificações de eventos e periódicos disponibilizados em um ritmo anual;

Objetivos específicos

1. Reestruturação da arquitetura e banco de dados do SILQ a fim de suportar classificações de eventos e periódicos disponibilizados em um ritmo anual;
2. Atualização do banco de dados do sistema com as últimas classificações disponibilizadas pelo Qualis (anos 2013 e 2014);

Objetivos específicos

1. Reestruturação da arquitetura e banco de dados do SILQ a fim de suportar classificações de eventos e periódicos disponibilizados em um ritmo anual;
2. Atualização do banco de dados do sistema com as últimas classificações disponibilizadas pelo Qualis (anos 2013 e 2014);
3. Criação de uma API pública de disponibilização dos serviços do SILQ, via camada de aplicação REST para integração com outros sistemas;

Objetivos específicos

4. Alterações na interface do sistema incluindo migração de *framework* de interface, inclusão de controles de *feedback*, novos gráficos de acompanhamento de grupos de pesquisa e melhorias gerais de usabilidade;

Objetivos específicos

4. Alterações na interface do sistema incluindo migração de *framework* de interface, inclusão de controles de *feedback*, novos gráficos de acompanhamento de grupos de pesquisa e melhorias gerais de usabilidade;
5. Propor novos algoritmos de avaliação baseados em similaridade textual e *feedback* de relevância e verificar se a taxa de acerto do sistema foi melhorada com tal ação.

Procedimentos metodológicos

1. Atualização tecnológica e arquitetural
2. Inclusão de *feedback* de relevância

Procedimentos metodológicos

1. Atualização tecnológica e arquitetural
 - ▶ Criação da camada *RESTful*;
 - ▶ Migração do *framework* de *interface*;
 - ▶ Alteração do modelo lógico p/ suporte aos novos dados Qualis;
2. Inclusão de *feedback* de relevância

Procedimentos metodológicos

1. Atualização tecnológica e arquitetural
 - ▶ Criação da camada *RESTful*;
 - ▶ Migração do *framework* de *interface*;
 - ▶ Alteração do modelo lógico p/ suporte aos novos dados Qualis;
2. Inclusão de *feedback* de relevância
 - ▶ Controles de captação de *feedback*;
 - ▶ Desenvolvimento do algoritmo de classificação baseado em *feedback*;
 - ▶ Avaliação experimental dos algoritmos.

Sumário

Introdução

Histórico e Justificativa

Objetivos

Conceitos

Information Retrieval e Data-matching

Desenvolvimento

Alterações tecnológicas

Uso de feedback de relevância

Conclusões

IR e Data Matching

- ▶ *Information Retrieval* (IR)
 - ▶ *query*
 - ▶ documentos
- ▶ *Data-Matching*
 - ▶ similaridade / dissimilaridade
 - ▶ *threshold*
- ▶ SILQ: sistema de IR baseado em *data matching*

n-grams / trigrams

- ▶ *n*-grams + trigrams
- ▶ qual threshold utilizar? qual função de similaridade utilizar?
- ▶ o sistema apresenta os resultados corretos?

Como o SILQ avalia um currículo Lattes

Artigo #1 (extraído do Lattes)

Título: Approximate data instance matching: a survey

Ano: 2011

Área: Ciência da Computação

Journal: Knowledge and Information Systems

ISSN: 0219-1377

Artigo #2 (extraído do Lattes)

...

Como o SILQ avalia um currículo Lattes

Artigo #1

Título: Approximate data instance matching: a survey

Ano: 2011

Área: Ciência da Computação

Journal: Knowledge and Information Systems

ISSN: 0219-1377

query: (ISSN, ano, área)

$q_A = (0219-1377, 2011, \text{Ciência da Computação})$

$q_A = (0219-1377, 2011, \text{Ciência da Computação})$

$q_A = (0219-1377, 2011, \text{Ciência da Computação})$

Conceito	Ano	ISSN	Título
A2	2011	0219-1377	Knowledge and Information Systems (Print)
A2	2012	0219-1377	Knowledge and Information Systems (Print)
B1	2014	0219-1377	Knowledge and Information Systems (Print)
A2	2010	0219-1377	Knowledge and Information Systems (Print)

Tabela: Resultados retornados pelo SILQ para a query q_A

$$q_A = (0219-1377, 2011, \text{Ciência da Computação})$$

Conceito	Ano	ISSN	Título
A2	2011	0219-1377	Knowledge and Information Systems (Print)
A2	2012	0219-1377	Knowledge and Information Systems (Print)
B1	2014	0219-1377	Knowledge and Information Systems (Print)
A2	2010	0219-1377	Knowledge and Information Systems (Print)

Tabela: Resultados retornados pelo SILQ para a query q_A

Resultado

Artigo #1 recebe o conceito A2

Como o SILQ avalia um currículo Lattes

Trabalho #1 (extraído do Lattes)

Título: A Strategy for Allowing Meaningful and Comparable Scores in Approximate Matching

Ano: 2007

Área: Ciência da Computação

Evento: Conference on Information and Knowledge Management (CIKM)

Como o SILQ avalia um currículo Lattes

Trabalho #1 (extraído do Lattes)

Título: A Strategy for Allowing Meaningful and Comparable Scores in Approximate Matching

Ano: 2007

Área: Ciência da Computação

Evento: Conference on Information and Knowledge Management (CIKM)

query: (título do evento, ano, área)

$q_T = (\text{Conference on Information and Knowledge Management (CIKM)"}, 2007, \text{Ciência da Computação})$

$q_T = (\text{Conference on Information and Knowledge Management (CIKM)}", 2007, \text{Ciência da Computação})$

$q_T = (\text{Conference on Information and Knowledge Management (CIKM)}", 2007, \text{Ciência da Computação})$

Conceito	Similaridade	Título
A1	0.71	International Conference on Information and Knowledge Management
B4	0.64	International Conference on Information, Process, and Knowledge Management

Tabela: Resultados retornados pelo SILQ para a query q_T

$q_T = (\text{Conference on Information and Knowledge Management (CIKM)}", 2007, \text{Ciência da Computação})$

Conceito	Similaridade	Título
A1	0.71	International Conference on Information and Knowledge Management
B4	0.64	International Conference on Information, Process, and Knowledge Management

Tabela: Resultados retornados pelo SILQ para a query q_T

Resultado

Trabalho #1 recebe o conceito A1

Métricas e avaliação de sistemas de IR

- ▶ Precisão e revocação
- ▶ taxa de acerto / exatidão
- ▶ conjunto de testes
- ▶ média de rank recíproco
- ▶ OFF: arquitetura?? (TODO)

Sumário

Introdução

Histórico e Justificativa

Objetivos

Conceitos

Information Retrieval e Data-matching

Desenvolvimento

Alterações tecnológicas

Uso de feedback de relevância

Conclusões

Obtenção de feedback

2016 | Ajustamento de pesos para ratings de múltiplos critérios em recomendação de itens
Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)

B3 23% 2012 WEBMEDIA Brazilian Symposium on Multimedia and the Web

2015 | Towards Automatic Document Classification by Exploiting only Knowledge Resources
International Conference of the Chilean Computer Science Society

B3 100% 2012 SCCC International Conference of the Chilean Computer Science Society

B2 64% 2012 ICSC_A International Computer Science Conference

[Ver menos resultados](#)

2015 | Implementação de um esquema de extração de dados tabulares da web
XII Workshop de Trabalhos de Iniciação Científica (WTIC)

Nenhum registro Qualis correspondente

Nenhum conceito encontrado | [Sugerir matching](#)

Sugerir matching

Nenhum registro Qualis correspondente

Figura: Controles de feedback da página de resultados de avaliação do SILQ

Algoritmo fb(t)

HAL! go ritmo

Algoritmo query_aliasing

HAL! go ritmo

Avaliação de *threshold* ideal

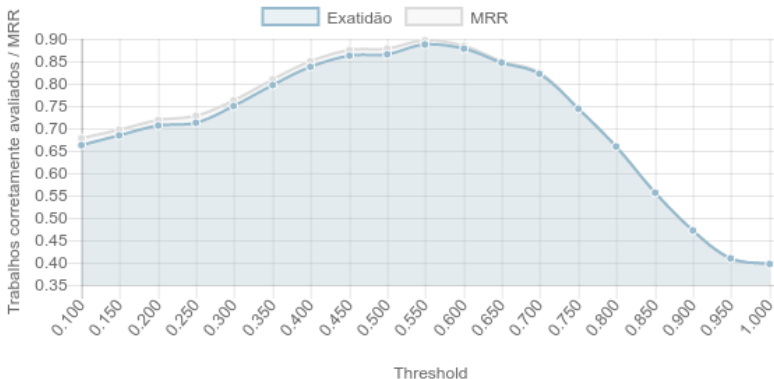


Figura: Valores de exatidão e MRR para diferentes valores de *threshold* utilizando o método *trigram*

Exatidão dos algoritmos propostos

Algoritmo	Exatidão
<i>trgm</i>	88.667%
<i>trgm + fb(1.00)</i>	89.667%
<i>trgm + fb(0.90)</i>	90.667%
<i>trgm + fb(0.80)</i>	92.667%
<i>trgm + fb(0.70)</i>	92.667%
<i>trgm + fb(0.60)</i>	91.000%
<i>trgm + query_aliasing</i>	93.333%

Tabela: Comparação da exatidão dos diferentes algoritmos testados (utilizando *threshold* de 0.55)

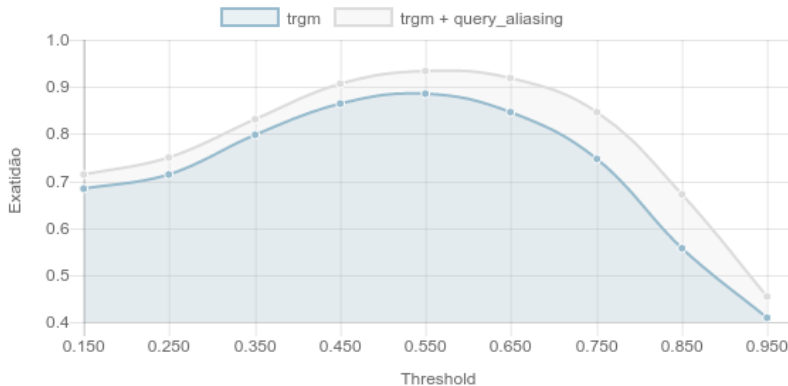


Figura: Comparação da taxa de acerto do algoritmo *trgm* e do *trgm + query_aliasing* para diferentes *thresholds*

Sumário

Introdução

Histórico e Justificativa

Objetivos

Conceitos

Information Retrieval e Data-matching

Desenvolvimento

Alterações tecnológicas

Uso de feedback de relevância

Conclusões

Conclusões

Conclusões.

Trabalhos futuros

Trabalhos futuros.

Análise do uso de *feedback* de relevância no Sistema de Integração Lattes-Qualis (SILQ)

Dúvidas?

Carlos Bonetti

carlosbonetti.mail@gmail.com