

Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- Método
- Resultados
- Conclusão



Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- Método
- Resultados
- Conclusão



Introdução







Somente a FAPESP faz a análise de 19 mil projetos de pesquisa por ano

38 mil horas gastas R\$3.8M*

Esse projeto busca construir uma ferramenta que auxilie em uma etapa do processo da pesquisa científica: a atribuição de avaliadores para que seja realizada a revisão por pares.



Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- Método
- Resultados
- Conclusão



Revisão por pares

PLN

Grafos

Emparelhamento

A revisão por pares é o procedimento utilizado no meio científico para avaliação de trabalhos, sejam eles artigos, projetos de pesquisa, etc.



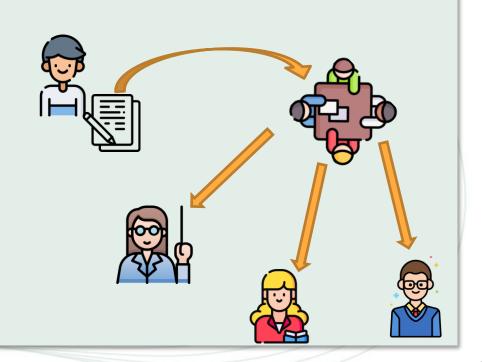
Revisão por pares

PLN

Grafos

Emparelhamento

A revisão por pares é o procedimento utilizado no meio científico para avaliação de trabalhos, sejam eles artigos, projetos de pesquisa, etc.





Revisão por pares

PLN

Grafos

Emparelhamento

Radicalização (stemming)

Técnica para extrair informação principal da palavra reduzindo-a ao seu radical

Saco-de-palavras (bag-of-words)

Técnica para simplificação de textos onde somente o conteúdos das palavras importa, sendo desconsiderados estrutura gramatical e ordem das palavras

Bigramas

São sequências de 2 palavras consecutivas em uma frase, utilizado para facilitar análise textuais

Estudo » Estud Estudante » Estud Estudar » Estud

"Como avaliar esse trabalho?"



"Como avaliar o como avaliar esse » o avaliar esse trabalho?" o esse trabalho



Revisão por pares

PLN

Grafos

Emparelhamento

Grafos são estruturas



Revisão por pares Emparelhamento PLN Grafos Ator Grafos são estruturas formados por vértices Ator Diretor Ator Diretor Ator Diretor Ator



Revisão por pares Emparelhamento PLN Grafos Ator Grafos são estruturas formados por vértices Ator Direto e arestas. Ator Diretor Ator Diretor Ator



Revisão por pares Emparelhamento PLN Grafos Ator Grafos são estruturas formados por vértices Ator Diretor e arestas. Pode ser um grafo Ator Diretor bipartido В Ator Direton Ator



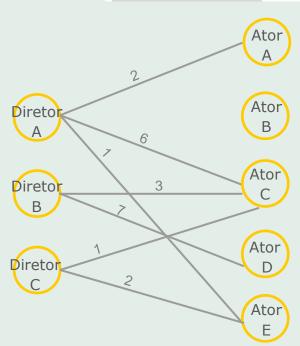
Revisão por pares

PLN

Grafos Emparelhamento

Grafos são estruturas formados por vértices e arestas.

Pode ser um grafo bipartido e ponderado.





Revisão por pares **PLN** Grafos Emparelhamento Emparelhamento é um conjunto de arestas de Ator Direto um grafo em que todos os vértices estão Ator Diretor conectados 2 a 2 e nenhuma aresta é Ator Diretor adjacente. Ator



Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- Método
- Resultados
- Conclusão



Método

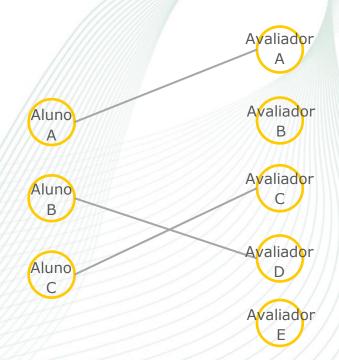
Avaliadores

- o Nome
- Artigos publicados
- Livros e capítulos de livros

Projetos

- Nome aluno
- o Título
- Palavraschave
- Resumo

 \Rightarrow



Input





Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,

Livro: Ensino de computação no ensino médio Artigo: Como o aprendizado de algoritmos auxilia no desenvolvimento cognitivo de adolescentes Palavras-chave: ensino, cognição



Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,

Remoção de stop words

Livro: Ensino
computação ensino
médio
Artigo:
aprendizado
algoritmos auxilia
desenvolvimento cognitivo
adolescentes
Palavras-chave: ensino,
cognição



Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,

Remoção de palavras frequentes

Livro:
computação
médio
Artigo:
aprendizado
auxilia desenvolvimento
cognitivo adolescentes
Palavras-chave:
Cognição



Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,

Remoção de acentos e pontuação

Livro:
computacao
medio
Artigo:
aprendizado
auxilia desenvolvimento
cognitivo adolescentes
Palavras-chave:
Cognicao



Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,

Stemming

Livro:
computac
med
Artigo:
Aprend
auxil desenvolvim
cogn adolescent
Palavras-chave:
Cogn



Exemplo

Prof. Doutor Bruno
Oliveira
Livro: Ensino de
computação no ensino
médio
Artigo: Como o
aprendizado de
algoritmos auxilia no
desenvolvimento cognitivo
de adolescentes
Palavras-chave: ensino,

Tratamentos textuais

Livro:
computac
med
Artigo:
Aprend
auxil desenvolv
cogn adolescent
Palavras-chave:
Cogn

Avaliador: Bruno Oliveira

desenvolv auxil med computac aprend adolescent

Avaliador: Bruno Oliveira

Cogn



Saco de palavras

desenvolv

auxil

med

computac

aprend

adolescent

Cogn

Avaliador: Bruno Oliveira



ID nó: Bruno Oliveira Bipartição: avaliador Saco-de-palavras: computac med Aprend

auxil desenvolv cogn

adolescent

















Avaliador E





cogn intel artif computac Avaliador

computac med aprend auxil desenvolv cogn adolescent

União 9 computac med

aprend auxil

desenvolv cogn adolescent intel artif

Intersecção 2 computac cogn

2/9 = 0,22

Coeficiente similaridade Aluno

Aluno В

Aluno

Avaliador

Avaliador В

Avaliador

Avaliador

Avaliador





cogn intel artif computac Avaliador

computac med aprend auxil desenvolv cogn adolescent

União 9 computac med

aprend auxil

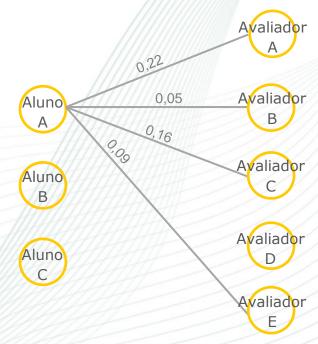
desenvolv cogn adolescent

intel

Intersecção 2 computac cogn

Coeficiente similaridade

2/9 = 0,22





artif



cogn intel artif computac Avaliador

computac med aprend auxil desenvolv cogn adolescent

União 9 computac med

aprend auxil

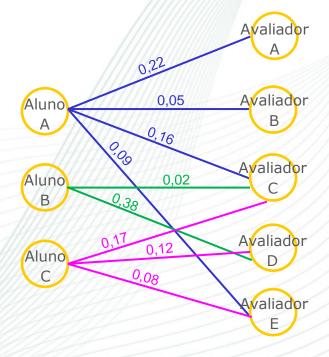
desenvolv cogn adolescent

intel artif

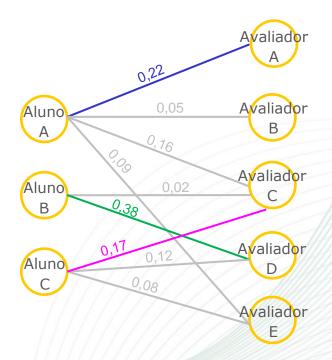
Intersecção 2 computac cogn

similaridade

Coeficiente 2/9 = 0,22







Algoritmo de Hopcroft-Karp

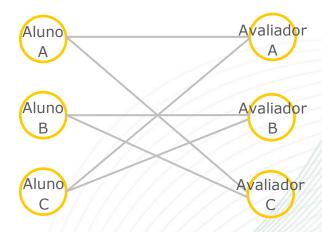
M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M



1ª iteração



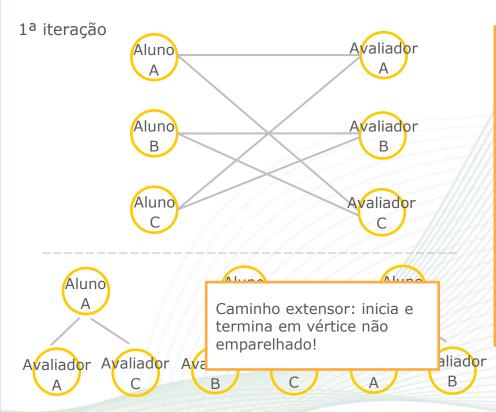
Algoritmo de Hopcroft-Karp

M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M





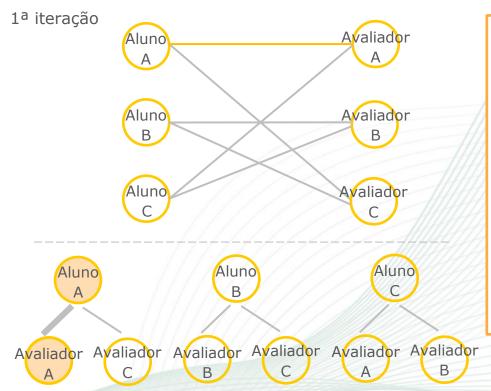
Algoritmo de Hopcroft-Karp

M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M





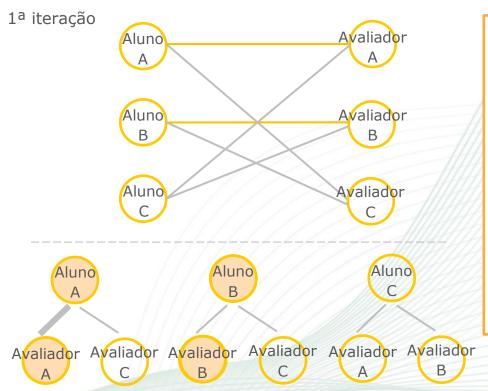
Algoritmo de Hopcroft-Karp

M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M





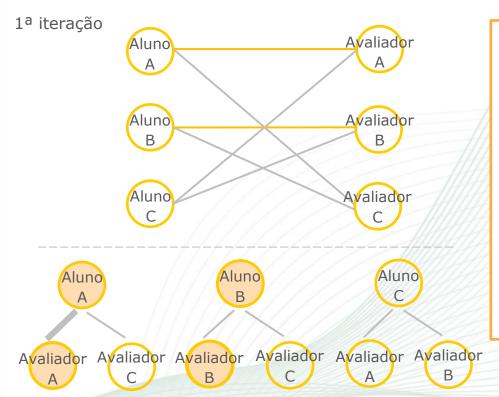
Algoritmo de Hopcroft-Karp

M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M





Algoritmo de Hopcroft-Karp

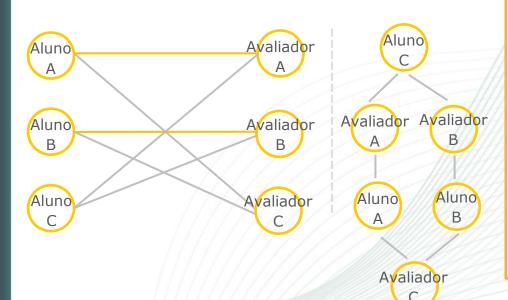
M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos **disjuntos** em G que aumentem o emparelhamento atual de M



2ª iteração



Algoritmo de Hopcroft-Karp

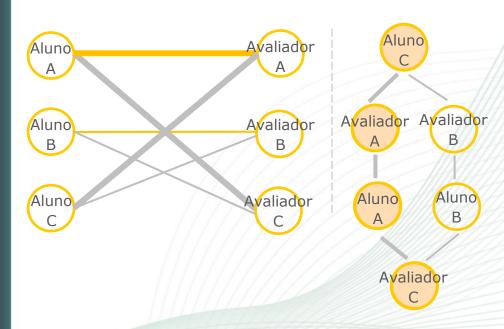
M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M



2ª iteração



Algoritmo de Hopcroft-Karp

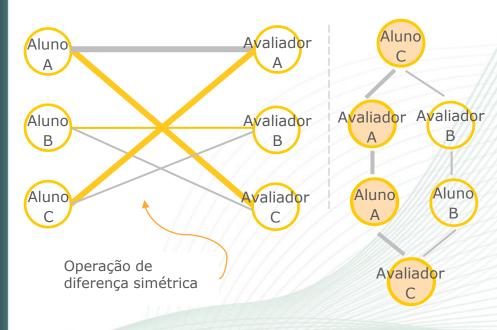
M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M



2ª iteração



Algoritmo de Hopcroft-Karp

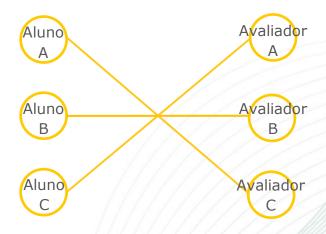
M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M



Método - Emparelhamento



Algoritmo de Hopcroft-Karp

M=0 enquanto houver caminho extensor para M

Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M



Método - Emparelhamento

Análise complexidade

Complexidade total:

$$O(|E|*\sqrt{|V|})$$



Algoritmo de Hopcroft-Karp

M=0 enquanto houver caminho extensor para M

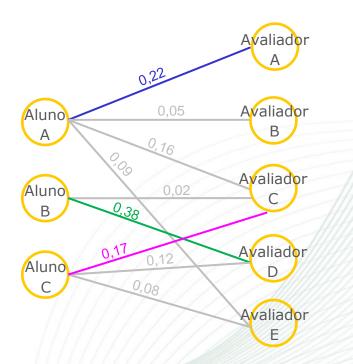
Usar uma BFS para construir um grafo G que se inicia nos vértices não emparelhados do subset A

Usar uma DFS para encontrar caminhos disjuntos em G que aumentem o emparelhamento atual de M

retornar M



Método – Emparelhamento



SUGESTÃO DE AVALIADORES PARA PROJETOS

Aluno A Avaliador A Aluno B Avaliador D Aluno C Avaliador C



Método - Complexidade do código

Imports e definições: 0(1)

Captura de dados: O(palavras_proj + palavras_aval)

Banco palavras: O(retirar)

Tratamento PLN: O((palavras_proj + palavras_aval)*retirar)

Criação dos vértices:

O(palavras_proj + palavras_aval)

Criação das arestas: O(palavras_proj * palavras_aval)

Emparelhamento: O(arestas*√(vértices))



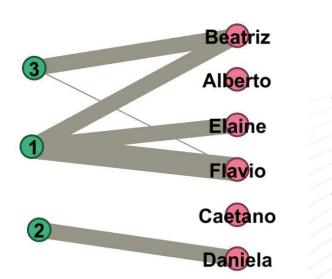
Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- Método
- Resultados
- Conclusão



Resultados – teste emparelhamento



Bipartição 1	Bipartição 2		
1	Flavio		
2	Daniela		
3	Beatriz		

O teste realizado apresentou o resultado esperado



Resultados – teste sem PLN

O teste realizado sem nenhum tratamento apresentou um número grande de conexões e altos coeficientes de similaridade

Claudio Fernando Andre Ana Carolina Boero

Lara Tenore Ferreira

Rafael Bruno Ferreira Figueiredo

Juliana Cretina Braga

Regina Helena de 🌒 iveira Lino Franchi

Vinícius Pazuch

Márcio Fablano da Silva

Alexandre Donizeti Alves

Marcia Aguiar

Cláudio Nei Nascimento da Silva

Amanda Crue Francesconi

Matheus Porto

Fellipe de Melo Chagas Roberto Patrus

Daniel Escudero Brito Silva

Jesus Pascual Mena Chalco

Itana Subiener

Lucia Regina Horta Rodrigues Franco

Vivili Maria Silva Gomes

Magda Aparecida Silverio Miyashiro Plínio Zernoff Táboas

Alessandro Jacques Ribeiro

Francisco José Brabo Bezerra

Evonir Albrecht

Claudia Josima Abrão de Araújo

Elisabete Marcon Mello

Pedro Paulo Amujo Maneschy

João Marcelo Sorovina Josko André Luiz Brandão

Virgínia Cardia Cardoso



Resultados – teste sem PLN

Ana Carolina Boero

Lara Tende Ferreira

Juliana Cristina Braga

Rataei Bruno Ferreira Figueiredo Regina Helena de Viveira Lino Franchi

O test tratar

Autor projeto	Avaliador sugerido	Peso	Termos em comum	Ĩ.
Amanda Cruz Francesconi	Virgínia Cardia Cardoso	0.076	uma, que, vez, a, estudo, em, é, um, para, com, o, no, da, de, projeto, ensino, como, na, processo, e, abordagem, número	3ilva \lves
Cláudio Nei Nascimento da Silva	Francisco José Brabo Bezerra	0.093	uma, dificuldades, que, A, a, estudo, os, da, partir, em, um, dos, com, o, no, sobre, as, de, onde, na, processo, do, e	Chalco
Daniel Escudero Brito Silva	Claudio Fernando Andre	0.072	em, de, como, estado, na, para, com, o, computacional, A, O, e, a, da, conhecimento	
Fellipe de Melo Chagas	Magda Aparecida Miyashiro	0.067	de, um, utilizando, para, o, a, da	mes
Lara Tenore Ferreira	Marcia Aguiar	0.084	uma, desenvolvido, que, história, desenvolvimento, A, a, em, partir, tarefa, O, para, com, o, no, da, de, como, contexto, na, das, do, e, por	Miyas pas tibeiro Bezer
Matheus Porto	Itana Stiubiener	0.075	em, uma, de, do, e, a, da	
Rafael Bruno Ferreira Figueiredo	fael Bruno Ferreira Figueiredo Lucia Regina Rodrigues Franco		uma, sistema, a, em, controle, O, um, dos, para, com, o, 3, da, as, rede, de, na, problema, do, e, aplicações	le Ara
Roberto Patrus	Plínio Zornoff Táboas	0.091	de, dos, para, na, História, o, no, e, sobre, da, as,	nesch

João Marcelo Borovina Josko

André Lue Brandão

Virginia Cardoso



Resultados – teste bigramas

O teste realizado com tratamento de PLN e utilizando bigramas apresentou um número baixo de conexões Lara Tende Ferreira

Robert®Patrus

Mathe Porto

Amanda Crue Francesconi Daniel Escudero Brito Silva

Cláudio Nei Nasimento da Silva Fellipe de Melo Chagas

Rafael Bruno Fereira Figueiredo

Evonir Albrecht
Plinio Zornoff Táboas
Jesús Pascual Mena Chalco

Francisco José Brabo Bezerra
Ana Carolina Boero
Silvia Cristina Dotta
Juliana Cristina Braga
Ronaldo Cristiano Prati
Diogo Santana Martins
Carlo Kleber da Silva Rodrigues
Vinicius Pazuch
Alexandre Donizeti Alves
Alexandre Balbinot
Yoss Zana

Debora Maria Rossi de Medeiros
João Marcelo Borovina Josko
Ailton Pauro de Oliveira Junior
Alessandro Jacques Ribeiro
Abrahão Fontes Baptista
Flávio Eduardo Aoki Horita
Carla Lopes Rodriguez
Joao Rioardo Sato

Raphael Yokoingawa de Camargo Sahudy Montenegro González Francisco Javist Ropero Peláez Rodrigo Pavão Itana Stubiener

> Marat @afikov Alexandre Mideki Okano

Edson Pinhero Pimentel



Resultados – teste bigramas

Lara Tendo-Ferreira

Robert Patrus

Juliana Cristina Braga

al Yokoingawa de Camargo

Sahudy Montenegro González

Francisco Javia Ropero Peláez

Rodrigo Pavão

Marat Rafikov

O tes trata

<u> </u>	///			Piogo Santana Martins
Autor projeto	Avaliador sugerido	Peso	Termos em comum	Kleber da Silva Rodrigues
Amanda Cruz Francesconi	Itana Stiubiener	0.007	estud uma	Vinicius Pazuch
Cláudio Nei Nascimento da Silva	Plínio Zornoff Táboas	0.008	as estud	Alexandr Balbinot
Daniel Escudero Brito Silva	Diogo Santana Martins	0.006	arte, estad	a Maria Rossi de Medeiros
Fellipe de Melo Chagas	João Marcelo Borovina Josko	0.008	de	Marcelo Borovina Josko
Lara Tenore Ferreira	Ana Carolina Boero	0.007	are uma, pesquis are	sandro Jacques Ribeiro
Matheus Porto	Evonir Albrecht	0.003	o conceit	rahão Fontes Baptista
Rafael Bruno Ferreira Figueiredo	Joao Ricardo Sato	0.000	conect red	arla Lope Rodriguez
Roberto Patrus	Debora Maria Rossi de Medeiros	0.005	management of	Joao Ridardo Sato
		-		a i i okolilyawa ue calilary

Cláudio Nei Nasimento da Silva

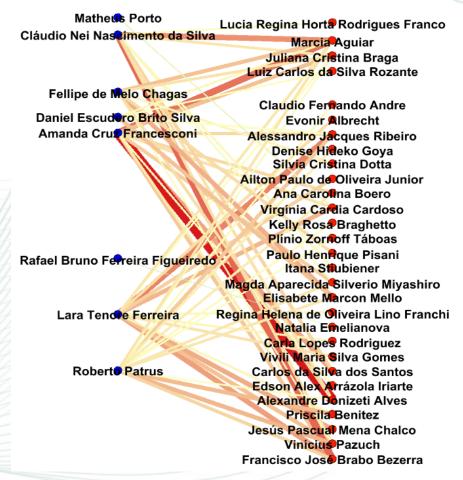
Fellipe de Melo Chagas

Rafael Bruno Fareira Figueiredo



Resultados – teste com PLN e palavras

O teste realizado com tratamento de PLN e utilizando somente palavras apresentou um número intermediário de conexões e coeficientes de similaridade





Resultados – teste com PLN e palavras

Matheds Porto
Cláudio Nei Nastimento da Silva

Lucia Regina Horta Rodrigues Franco

Marcia Aguiar Juliana Cristina Braga Luiz Carlos da Silva Rozante

Fellipe de Melo Chagas

Daniel Escudero Brito Silva

Claudio Fernando Andre Evonir Albrecht

O test tratan somer um nú conex simila

utor projeto Avaliador sugerido		Peso	Termos em comum	cques Ribeiro	
Amanda Cruz Francesconi	Alexandre Donizeti Alves	0.079	especif, bas, linguag, estud, are, ferrament, natural,	eko Goya	
Amanda Cruz Francesconi	Alexandre Donizett Aives	0.079	process	ina Dotta Oliveira Junior	
Cláudio Nei Nascimento da Silva	Francisco José Brabo Bezerra	0.065	exist, conhec, part, estud, signific, onde, compar,	na Boero	
Claudio Nei Nascimento da Silva	Francisco Jose Brado Bezerra	0.065	dificuldad, process	lia Cardoso	
Daniel Escudero Brito Silva	Elisabete Marcon Mello	0.050	especial, constru, propost, registr	Braghetto	
Fellipe de Melo Chagas	Marcia Aguiar	0.050	curricul, bas, pesquis, conhec, process	off Táboas	
			calcul, compreensa, histor, divers, futur, busc, conjunt,	que Pisani	
Lara Tenore Ferreira	Alessandro Jacques Ribeiro	0.061	signific, context, projet, impact, pesquis, consider, part,	ubiener	
			estud, taref, ii, relaco, process	Silverio Miyashiro arcon Mello	
Matheus Porto	Virgínia Cardia Cardoso	0.037	filosof, profund, leitur, critic, construca	liveira Lino Franc	
Rafael Bruno Ferreira Figueiredo	Natalia Emelianova	0.034	encontr, pont, simpl, font, red, uso, ambient	relianova	
Dalama Datuma	Vinícius Pazuch	0.052	pesquis, of, histor, review, cientif, uso, teoric, artig,	Rodriguez	
Roberto Patrus	Vinicius Pazucn	0.052	process	ilva Gomes	

a dos Santos
Edson Alex Arrázola Iriarte
Alexandre Donizeti Alves
Priscila Benitez
Jesús Pascual Mena Chalco

Francisco José Brabo Bezerra



Resultados – teste com dados reais

O teste realizado para uma aplicação real de atribuição considerou todo o tratamento de PLN e o uso de somente palavras. Os ajustes feitos em relação aos resultados anteriores:

- Três avaliadores por projeto
- Máximo de dois projetos por avaliador

Candidato	Avaliador 1	Similaridade	Termos em comum	Avaliador 2	Similaridade	Termos em comum	Avaliador 3	Similaridade	Termos em comum
1	Α	0.1267	118	E	0.1128	94	A C	0.1116	125
2	В	0.0859	70	AL	0.054	65	Υ	0.0988	131
3	С	0.0682	121	A M	0.0997	158	G	0.1236	203
4	D	0.095	62	L	0.0853	49	A B	0.0647	29
5	E	0.0947	76	Α	0.0965	88	0	0.0787	143
6	F	0.0702	71	R	0.1239	163	AQ	0.0918	147
7	G	0.1208	186	P	0.0518	69	1	0.0939	131
8	Н	0.0947	82	АН	0.1117	79	ΑH	0.1119	79
9	f f	0.0908	114	Х	0.0642	76	-	-	
10	J	0.0516	132	С	0.0595	115	AA	0.0634	95



Seleção de avaliadores de projetos: uma abordagem baseada em PLN e algoritmos de emparelhamento máximo

Sumário

- Introdução
- Conceitos básicos
- Método
- Resultados
- Conclusão



Conclusão

Melhor resultado

Grafo gerado com todas as técnicas de PLN e palavras ao invés de bigramas apresentou resultado satisfatório para o objetivo proposto com boa correlação entre avaliadores e projetos de pesquisa.

Ao ser testado em uma aplicação real de atribuição de avaliadores conseguiu auxiliar na tomada de decisão do órgão responsável.

Possíveis melhorias

- O Input manuais para ajustes finos
- Software para facilitar o acesso a ferramenta

Contribuição

Método inusitado na literatura para resolução de problema ainda em aberto e que ocorre em abundância na atualidade.



Referências

- Bond, J. A., & Murty, U.S. 1976. Graph theory with applications. The Macmillan Press Ltd.
- Digiampietri, Luciano Antonio, Mena-Chalco, Jesus Pascual, Perez-Alcazar, Jose de Jesus, Fernandez Tuesta, Esteban, Delgado, Karina Valdivia, Mugnaini, Rogerio, & Silva, Gabriela S. 2012. Minerando e caracterizando dados de curriculos lattes. Pages 1–12 of: Brazilian workshop on social network analysis and mining – brasnam
- Erciyes, Kayhan. 2015. Complex networks: An algorithmic perspective. CRC Press.
- Fernando A. Ribeiro Serra, Manuel Portugal Ferreira, Gabriela Gon calves Fiates Lenise Costa. 2007. Publicar e dificil ou faltam competencias? o desafio de pesquisar e
 publicar em revistas científicas na visao de editores e revisores internacionais. Encontro
 de ensino e pesquisa em administracao e contabilidade, 9(4).
- Jurafsky, Daniel, & Martin, James. 2000. Speech and language processing. Prentice Hall.





Obrigada!

