

Palavras-chave: Bases de dados; Interoperabilidade; IRaMuTeQ.

Keywords: Data bases; Interoperability; IRaMuTeQ.

1. Introdução

Diante do aumento crescente da digitalização de documentos e produção digital de conhecimento em diversos tipos e formatos de dados, a questão da interoperabilidade tornou-se um problema emergente nas áreas de Ciência e Tecnologia. A recuperação da informação não se restringe mais a unir termos pré-indexados a documentos, mas também procura resolver problemas conceituais e semânticos que envolvem a correlação entre termos, palavras-chave e conceitos para diferentes objetos de dados.

O presente artigo visa levantar a discussão sobre a interoperabilidade mediante o relato de uma pesquisa empírica-descritiva realizada nas bases de dados Scopus, LISA (Library and Information Science Abstracts) e Web of Science a partir dos termos *Amazônia* e *Amazon*, por meio da comparação de distâncias semânticas entre termos e conceitos, através do resultado da aplicação do programa IRaMuTeQ.

2. Metodologia

A pesquisa de caráter empírico utilizou como base a teoria do conceito de Dahlberg (2014), a análise de domínio de Hjørland (2002) e os princípios FAIR (WILKINSON *et al.*, 2016), aplicados à análise de similitude (teoria dos grafos) do IRaMuTeQ, uma ferramenta de análises textuais estatísticas ancorada às linguagens R e Python (CAMARGO e JUSTO, 2013).

Foram analisados 120 resumos de artigos recuperados nas bases de dados Scopus, LISA e Web of Science, no período de 2009 a 2018, através dos termos *Amazônia* e *Amazon*, reunidos em *corpora* textuais por base de dados para análise dos resultados.

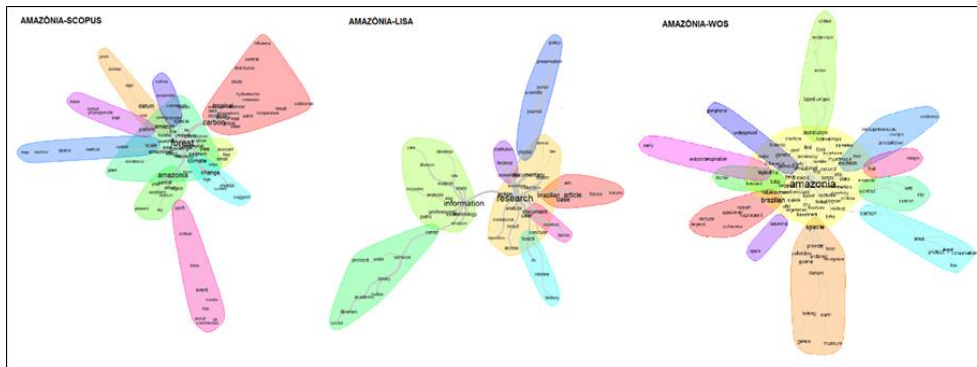
3. Resultados por termo e base de dados

Na Scopus, a análise da similitude de *Amazônia* mostrou o termo *forest* representando o *cluster* central e com atração semântica com os seguintes termos: *large*, *datum*, *amazon*, *tree*, *climate*, *change*, *carbon*, *tropical*, *amazônia*. No grafo existe uma convergência dos *subclusters* para o *cluster* do termo *forest*, sendo o *cluster* do *uplift* o mais distante e ligado pelo *subcluster* do termo *amazon*.

Na LISA, a similitude de *Amazônia* apresentou um *cluster* principal identificado pelo termo *research*, formado por cinco *subclusters* representados pelos termos: *brazil*, *document*, *brazilian*, *document* e *digital*. No *cluster information* foi identificado o termo *amazon* e este está ligado ao *subcluster center*.

Na Web of Science (WOS), a similitude de *Amazônia* apontou um *cluster* principal simbolizado pelo termo *amazônia* composto por dez *subclusters* identificados pelos termos: *specie*, *savanna*, *brazilian*, *total*, *highmodinia*, *wester*, *distribution*, *eastern*, *fruit* e *central*. Além disso, o *cluster* central está ligado ao *cluster carbon*. Convém mencionar que o termo *amazon* apareceu próximo aos *subclusters central e fruit*.

Fig. 1 – Análise de similitude de Amazônia na Scopus, LISA e WoS



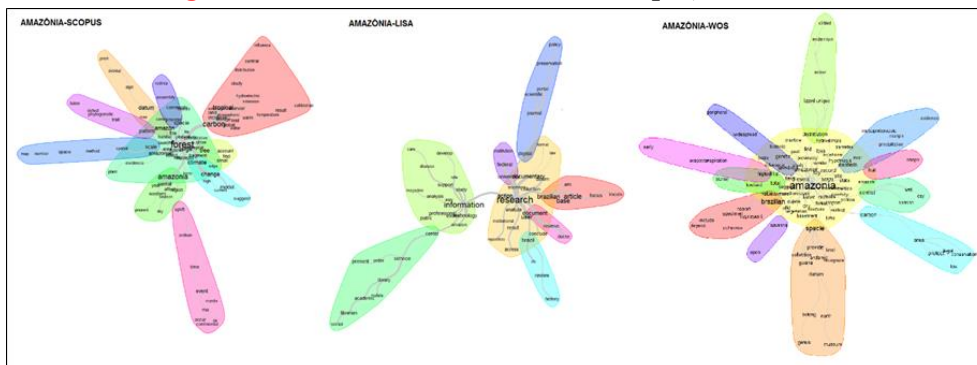
Fonte: Elaborado pelos autores com base no resultado do IramuTeQ.

Na Scopus, a análise da similitude de *Amazon* não apontou um termo central. Os *clusters* são dispersos e organizados nas seguintes sintaxes: 1) na posição logo acima do *cluster amazon* estão os *clusters* dos termos: *mechanical*, *turk*, *amazon* e *show*; 2) na posição à direita do *cluster amazon* estão os *clusters* dos termos: *online*, *research*, *review*, *product* e *datum*; 3) na posição superior estão os *clusters* dos termos: *mturk*, *datum* e *system*; 4) junto ao termo *amazon* foram encontrados termos referentes à floresta como: *soil*, *forest* e *carbon*.

Na LISA, a similitude de *Amazon* exibiu um núcleo central representado pelo termo *amazon* conectado a três *subclusters* identificados pelos termos: *book*, *online* e *mechanical*, em um aparente movimento de dispersão. Não foram encontrados termos referentes à floresta.

Na WoS a similitude de *Amazon* mostrou o termo *amazon* simbolizando o *cluster* principal, o qual está conectado a dez *subclusters* representados pelos termos: *low*, *strategy*, *brazilian*, *system*, *time*, *development*, *datum*, *model*, *develop* e *forest* (com *subcluster deforestation*), assim como está vinculado ao *cluster* do termo *support*.

Fig. 2 – Análise de similitude de Amazon Scopus, LISA e WoS



Fonte: Elaborado pelos autores com base no resultado do IramuTeQ.

4. Considerações

O resultado da análise de similitude revelou a não interoperabilidade dos termos *Amazônia* e *Amazon* quanto às suas relações de complementaridade, oposição e relações partidárias, visto que o resultado da busca pelo termo *Amazon* recuperou tanto domínios referentes à floresta quanto sobre a empresa de tecnologia Amazon. Para alguns usuários de bases de dados, os termos *Amazônia* ou *Amazon* seriam os mesmos, ou seja, simplesmente uma questão de tradução. Já para outros, estes termos podem traduzir diferentes sistemas de valores culturais.

Referências bibliográficas

CAMARGO, Brígido V.; JUSTO, A. M.

2013 IRAMUTEQ: um software gratuito para análise de dados textuais. *Temas em Psicologia*. 21:2 (2013) 513-518.

DAHLBERG, I.

2014 What is knowledge organization? *Knowledge Organization*. 41:1 (2014) 85-91.

HJØRLAND, B.

2002 Domain analysis in Information science: eleven approaches, traditional as well as innovative. *Journal of Documentation*. 58 :4 (2002) 447-448.

WILKINSON, M. D. [et al.]

2016 The FAIR guiding principles for scientific data management and stewardship. *Nature: Scientific Data* (2016).

Marcos Gonçalves Ramos | marcosgon54@hotmail.com

PPGCI IBICT/ECO-UFRJ, Brasil

Priscila Ramos-Carvalho | carvalho.priscila@gmail.com

PPGCI IBICT/ECO-UFRJ, Brasil

Fábio Castro Gouveia | fgouveia@gmail.com

Fundação Oswaldo Cruz, Brasil