

# PALAVRAS-CHAVE COMO ELO ENTRE ARTIGOS E AUTORES: visualizações possíveis

Elaine Oliveira Lucas (UDESC)

[lanil@udesc.br](mailto:lanil@udesc.br)

Marilda Lopes Ginez Lara (USP)

[larama@usp.br](mailto:larama@usp.br)

EIXO TEMÁTICO: Mapas da Ciência

MODALIDADE: Apresentação oral

## 1 INTRODUÇÃO E OBJETIVOS

Integrando uma pesquisa mais ampla, este trabalho se propõe a explorar alguns dos elementos que permitiriam contribuir para traçar um mapa de tendências das pesquisas sobre Knowledge Organization (KO), focalizando especificamente, para o período compreendido entre 1997 e 2011, nos idiomas francês, espanhol e português, a incidência de artigos relativos à KO que abordam questões sobre Linguagens Documentárias e Sistemas de Organização do Conhecimento.

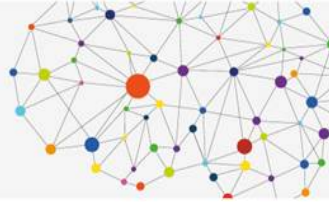
O objetivo geral desta parte da pesquisa foi o de verificar, a partir de uma equação de busca e de um *corpus* pré-definidos, que palavras-chave são privilegiadas e que redes científicas podem ser identificadas relacionando estas palavras-chave aos autores dos artigos. Como objetivo específico, pretendemos apresentar graficamente os resultados das relações estabelecidas entre os artigos, por meio da identificação de coocorrência de palavras-chave e, entre autores e palavras-chave, de modo a oferecer um quadro para uma melhor percepção das redes científicas constituídas, redes essas normalmente não evidenciadas por outros estudos métricos.

Partimos do pressuposto que estudos dessa natureza constituem uma ferramenta útil para a identificação das tendências de pesquisa, cuja análise pode trazer contribuições importantes para o aperfeiçoamento de metodologias relacionadas ao tratamento da informação, com consequências para o ensino, a pesquisa e a interlocução entre pesquisadores.

## 2 OPÇÕES METODOLÓGICAS

O *corpus* da pesquisa foi definido a partir da equação de busca abaixo:

*“Bibliographic classification” OR Categories OR Categorization OR Classification OR “Classification of Knowledge” OR “Classification scheme” OR “Classification system” OR “Controlled language” OR “Controlled vocabulary” OR “Document analysis” OR “Document organization” OR “Documentary language” OR “Facet analysis” OR “Facet classification” OR*



“Facet classification schemes” OR “Index languages” OR Indexing OR “Indexing language” OR “Information organization” OR “Knowledge Organization” OR “Knowledge Representation” OR “Knowlege Organization Systems” OR KOS OR “Markup language” OR “Natural languages” OR Ontologies OR Ontology OR “Retrieval Languages” OR “Semantic vocabularies” OR “Simple Knowledge Organization System” OR SKOS OR “Strutured vocabulary” OR “Subject analysis” OR “Subject approach to information” OR “Subject cataloguing” OR “Subject heading” OR “Subject Headings Schemes” OR “Subject Indexes” OR “Subject Indexing” OR “Synonym rings” OR Taxonomies OR Terminologies OR Thesauri OR Thesaurus OR “Topic Maps” OR Vocabularies OR “Vocabulary control” OR “web semantics”) AND (Language OR “Language theories” OR Linguistics OR “Linguistics theories” OR Semantics OR Terminology OR Lexicology)

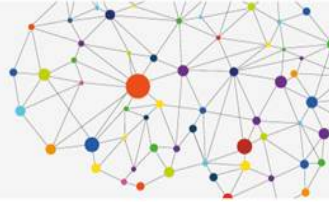
A delimitação temporal da pesquisa compreendeu a literatura publicada entre 1997 e 2011, em espanhol, francês e português, repertoriada pelas bases de dados *Library Information Science Abstract* (LISA), *Library, Information Science & Technology Abstracts* (LISTA), *Scopus* e *Web of Science* (WoS). A mesma equação de busca foi utilizada nas quatro bases de dados, mantendo-se as palavras-chave em inglês utilizadas por cada uma das bases de dados, mesmo quando o idioma original dos artigos fosse diferente.

Para a análise do *corpus* foram utilizados alguns *softwares*, de acordo com as diferentes fases da pesquisa. Foi utilizado o *Microsoft office Excel 2013*® para separar e minerar os dados; o *Software R*® foi adotado para desenvolver os pares relacionais (matrizes) necessários para o desenvolvimentos dos grafos; e, em seguida, para o desenvolvimento dos grafos com atributos e métricas de acordo com as relações estabelecidas, o *software* livre *Gephi*® versão 0.8.2 – Beta.

### 3 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

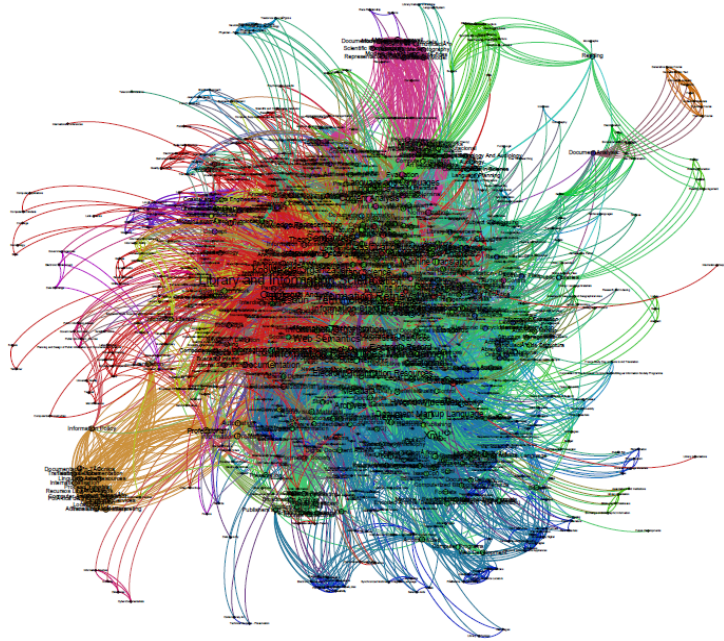
Dos 354 artigos que formam o *corpus* desta pesquisa foram analisadas todas as palavras-chave recuperadas. A maioria delas está no idioma inglês e, em alguns casos, alguns termos nos idiomas do artigo original também foram recuperados.

A expansão da equação de busca, com a inclusão dos termos de indexação utilizados pelas bases de dados, resultou na identificação de 927 palavras-chave diferentes nos 354 artigos escritos por 462 autores. Estas palavras-chave foram utilizadas 2.611 vezes. No entanto, 648 delas foram utilizadas apenas uma vez, inviabilizando, obviamente, a geração de uma rede de artigos.



### 3.1 COCORRÊNCIA DE PALAVRAS-CHAVE

O Grafo 01 apresenta todas as relações de coocorrência de palavras-chave existentes no *corpus* analisado. Alguns agrupamentos destas coocorrências foram destacados, pelo *software* Gephi®, por diferentes cores.



Grafo 01 – Rede de relações entre palavras-chave – rede total

A visualização do grafo acima, por seu tamanho – 22.839 pares de coocorrência entre as palavras-chave – fica prejudicada. No entanto, limitando as palavras-chave com uma quantidade de relações igual ou superior a 50, chegamos ao grafo abaixo, onde é possível visualizar somente os laços entre as palavras-chave de maior incidência relacional.

O Grafo 02 também permite visualizar as palavras-chave que possuem maior quantidade de relações com outras e, portanto, as relações entre os pares de palavras-chave mais utilizados para representar o *corpus*.









que é detentora de forte poder simbólico e de autoridade no campo, o que lhes concede o privilégio de definir as várias questões na estrutura do campo.

Além disso, as relações entre palavras-chave e autoria demonstram o surgimento e a possibilidade de análise de comunidades ligadas por afinidade terminológica na representação de sua produção no campo. As diferenças terminológicas estão relacionadas “aos níveis de abordagem da representação da informação ou ao foco privilegiado: objeto de análise, processos, produtos, instrumentos de representação” (LARA, 2011, p. 104, citada por LUCAS, 2014) dos artigos que compuseram o *corpus*.

#### 4 CONSIDERAÇÕES FINAIS

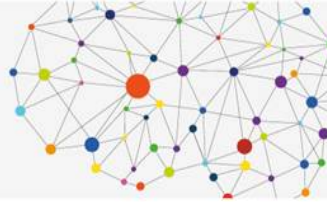
A pesquisa permitiu recuperar artigos que abordassem questões sobre as Linguagens Documentárias e os Sistemas de Organização da Informação formalmente identificadas à KO ou que lhe fossem próximas, considerando a produção científica nos idiomas citados e repertoriada por bases de dados internacionais no período compreendido entre 1997 e 2011.

Foram adotados diferentes métodos de pesquisa para alcançar os objetivos traçados. A combinação dos métodos adotados permitiram compreender melhor a dinâmica da atividade científica do campo metodologicamente desenhado para esta investigação.

Em síntese, foram utilizadas 927 palavras-chave para descrever os 354 artigos que compuseram o *corpus*, das quais 648 palavras foram utilizadas apenas uma vez. A análise de coocorrência entre palavras-chave recuperou 22.839 pares de palavras-chave. Considerando apenas a incidência de palavras-chave com quantidade de relações igual ou superior a 50, foram obtidas 75 palavras-chave.

As coocorrências de palavras-chave com maior incidência (acima de 30) encontradas no *corpus* foram: ‘*Web Semantics – Ontology*’; ‘*Automatic Indexing - Automatic Text Analysis*’; ‘*Automatic Indexing - Machine Translation*’; ‘*Automatic Text Analysis - Machine Translation*’; ‘*Ontology - Library and Information Science*’; ‘*Terminology - Library and Information Science*’;

A rede de relações entre palavras-chave e autores foi apresentada de forma completa (Grafo 03) e, em seguida, com destaque para as relações com grau igual ou superior a 20 (Grafo 04) e enfoque na formação de comunidades (Grafo 05).



Concluindo, podemos afirmar que a validade de uma pesquisa tem de ser considerada em relação ao universo delimitado e às suas metodologias. Isso significa dizer que não é possível, a partir dos resultados, realizar generalizações. Os recortes operacionalizados e as opções metodológicas para sua observação impõem limitações, mas são ferramentas importantes para a análise de problemas específicos. Podemos afirmar, todavia, que o modelo metodológico aplicado se prestaria ao estudo de vários aspectos da produção científica. Sua adaptação a diferentes realidades, e/ou campos, dependeria da introdução de outras variáveis que atuam sobre as interações, conforme o interesse do estudo.

## REFERÊNCIAS

LABORATÓRIO de Estudos Métricos da Informação (LEMI). Disponível em: <http://lemi.uc3m.es/> Acesso em 28 fev. 2014.

LUCAS, Elaine Rosangela de Oliveira. **Capital Social e Capital Científico na produção científica sobre Linguagens Documentárias e Sistemas de Organização do Conhecimento no campo da Knowledge Organization (KO) nos idiomas espanhol, francês e português.** São Paulo, 2014. 165f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2014.

SOFTWARE R®. Disponível em: <http://www.r-project.org/> Acesso em 28 fev. 2014.

SOFTWARE livre Gephi® versão 0.8.2 – Beta. Disponível em: <https://gephi.org/> Acesso em 28 fev. 2014.