



Mapping Terms and Entities with Hypergraphs

Creative Collab 2019

Last update: June 18, 2019

José Devezas

MAP-i 2016/2017

INESC TEC & FEUP InfoLab

Supervisor:

Sérgio Nunes

U.Porto

External advisor:

Bruno Martins

U.Lisboa



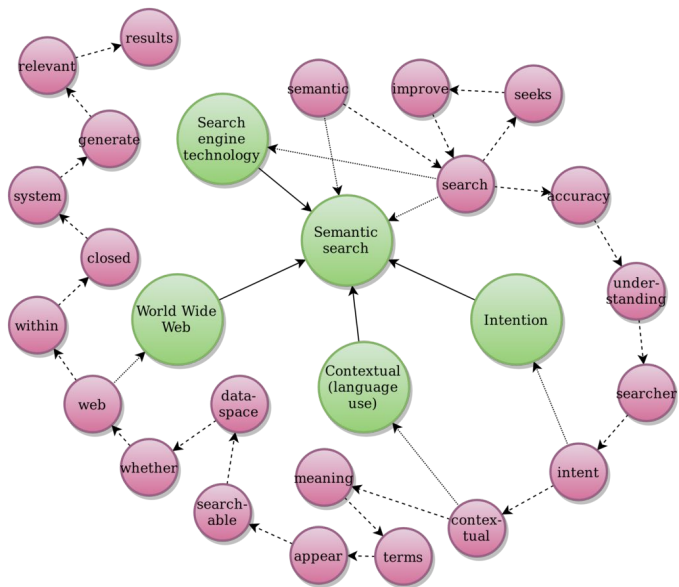
Universidade do Minho



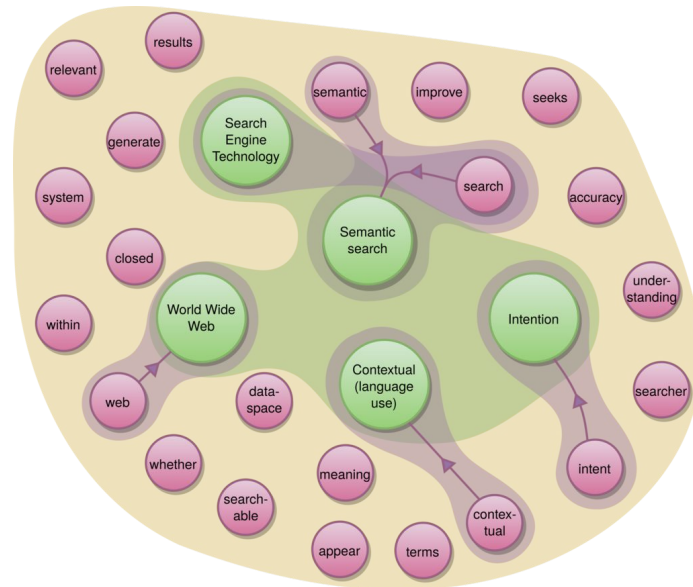


Hypergraph

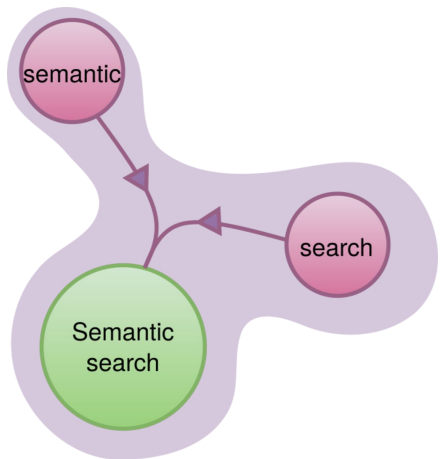
- A graph models dyadic relations between two elements.
- A hypergraph is a generalization of a graph, where edges (or rather hyperedges) are able to establish n -ary relations between an arbitrary number of elements.
- Hyperedges can be:
 - Undirected (a set of nodes) $\{n_1, n_2, n_3\}$
 - Or directed (a pair of sets of nodes) $(\{n_1, n_2\}, \{n_3\})$



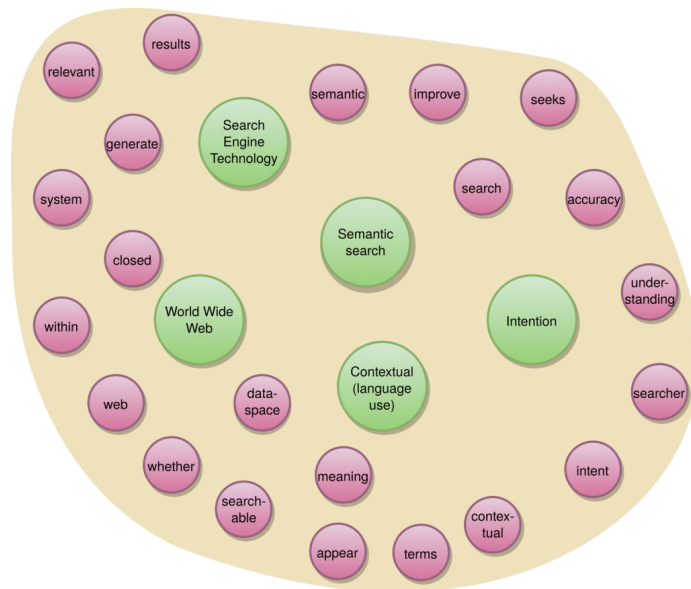
Graph: dyadic/binary relations.



Hypergraph: n -ary relations.



Directed hyperedges
(e.g., contained in).



Undirected hyperedges
(e.g., a bag of words and entities).

**What can we represent with
hypergraphs? Are there
limitations?**

—



Hypergraph expressiveness

Strengths:

- n -ary relations.
- Hierarchical relations.
- Overlapping relations.
- Weighted nodes and hyperedges.
- Directed dependencies between two groups.

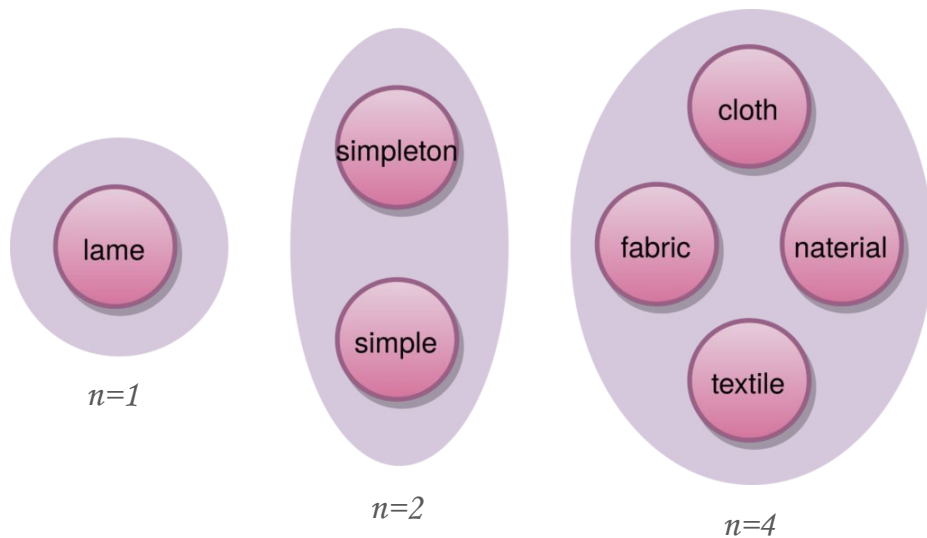
Weaknesses:

- No membership degree (i.e., a node either belongs to a hyperedge or it doesn't).
- Non-explicit dependencies between groups.

Strengths

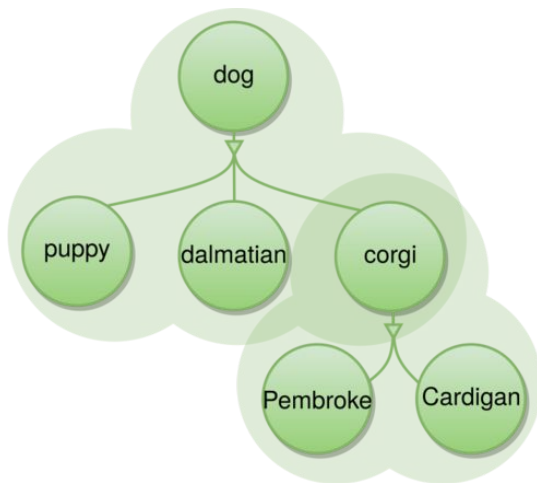
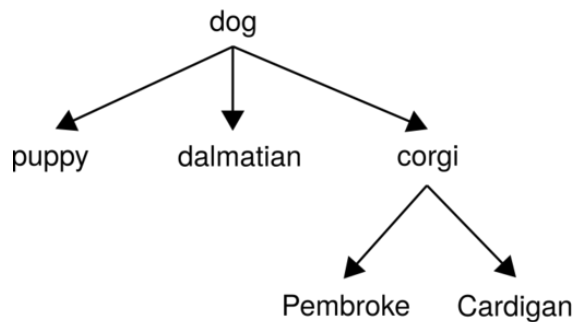


n -ary relations



Hyperedge cardinality: Number of nodes n that belong to a hyperedge.

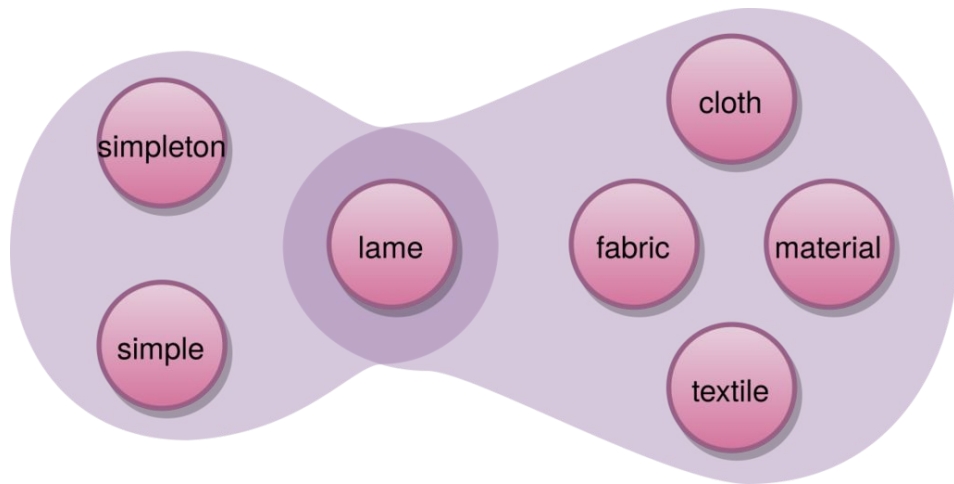
Hierarchical relations



e.g., hyponyms: Two levels of hyponymy, starting from 'dog', and then from 'corgi'.



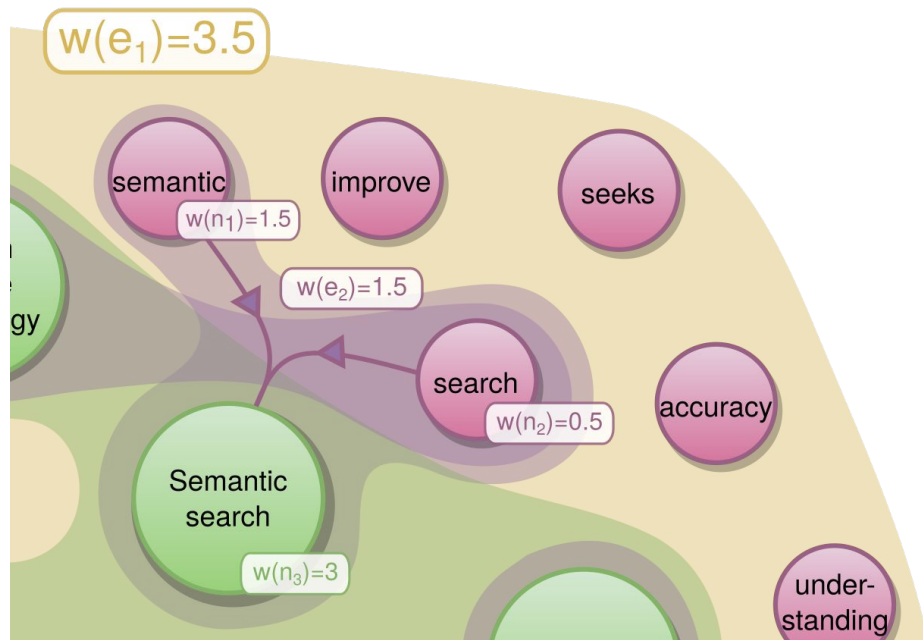
Overlapping relations



e.g., **synonyms**: 'lame' has two senses that can be represented by two overlapping undirected hyperedges.

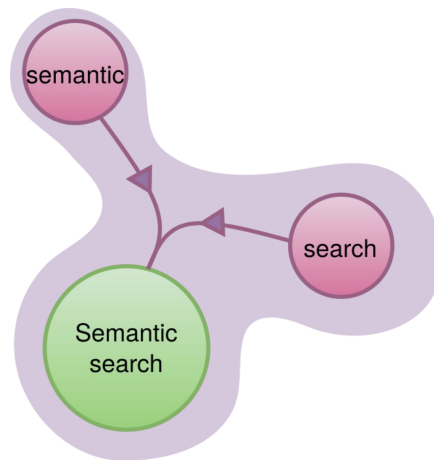
Weighted nodes and hyperedges

Weighted hypergraph: different degrees of importance can be assigned to nodes and hyperedges. These are globally set for the hypergraph.



Directed dependencies between two groups

- While we still have an n -ary relation...
- ...The dependency is binary.
- It's between a group of two terms ('semantic' and 'search')...
- ...And a group of one entity ('Semantic search').



$(\{\textit{semantic}, \textit{search}\}, \{\textit{Semantic search}\})$

Weaknesses

...and how to mitigate them.

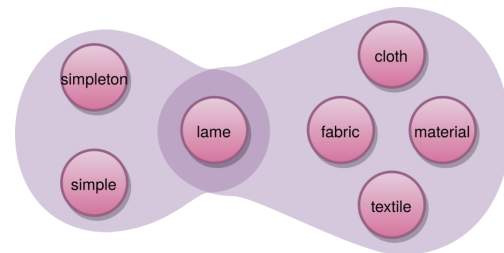
No membership degree

...solved through fuzzy hypergraphs.

Undirected hyperedges:

$\{lame, simpleton, simple\}$

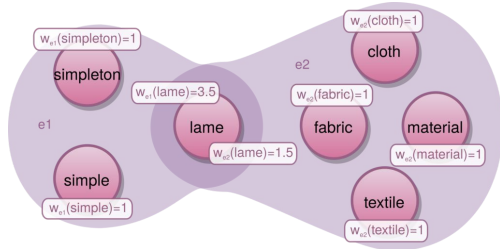
$\{lame, fabric, cloth, material, textile\}$



- A node either belongs to a hyperedge or it doesn't.
- In order to set a membership degree, we need a fuzzy hypergraph (Lee-Kwang and Lee, 1995).

No membership degree

...solved through fuzzy hypergraphs.



Undirected fuzzy hyperedges:

$$\{(lame, w(lame) = 3.5), \\ (simpleton, w(simpleton) = 1), \\ (simple, w(simple) = 1)\}$$

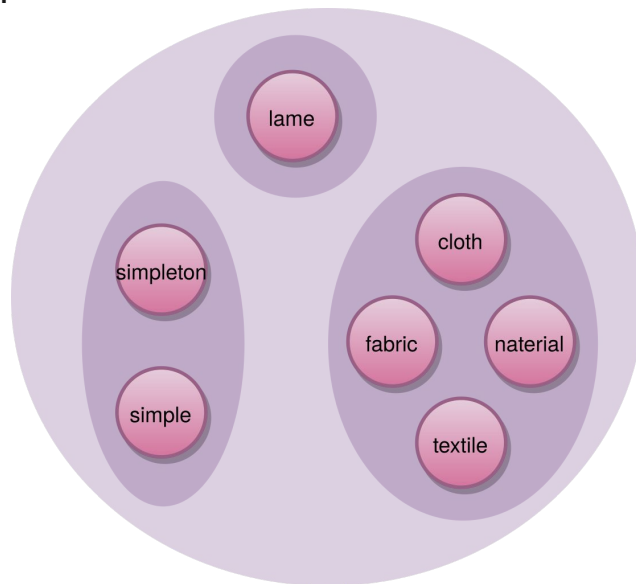
$$\{(lame, w(lame) = 1.5), \\ (fabric, w(fabric) = 1), \\ (cloth, w(cloth) = 1), \\ (material, w(material) = 1), \\ (textile, w(textile) = 1)\}$$

- Where instead of defining a hyperedge as a set of nodes...
- ...We define a fuzzy hyperedge as a set of pairs of node and weight function.

Non-explicit dependencies between groups

...mitigated through hierarchical combinations of hyperedges.

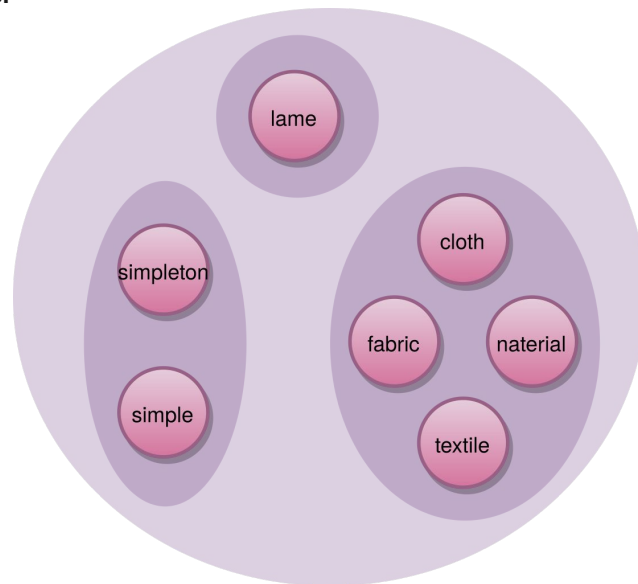
- Directed hyperedges are able to establish dependencies between two groups of nodes.
- But how do we model the dependencies between $n > 2$ groups?



Non-explicit dependencies between groups

...mitigated through hierarchical combinations of hyperedges.

- One way would be to define a hyperedge that contains the n hyperedges for the groups.
- Although a hyperedge cannot directly contain other hyperedges, it can contain their nodes.
- Subsumption is then verifiable rather than explicitly represented.



**Has anyone used hypergraphs
for the representation or
retrieval of documents?**

—



Topics maps

- One of the first hypergraph-based models for indexing documents.
- Built to merge back-of-the-book indices.
- And to improve findability.
- In practice, though, few work exists on topics maps applied to search.

Nodes:

- **Topics** (e.g., entities, types);
- **Occurrences** (information resources, e.g., document identifier+page).

Edges:

- Binary connections between topics and occurrences;
- **Associations:** *n*-ary connections between topics.

Bendersky, M., & Croft, W. B. (2012, August). Modeling higher-order term dependencies in information retrieval using query hypergraphs. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (pp. 941-950). ACM.



Query hypergraph

Nodes:

- **Concept** (terms, bigrams, noun phrases, named entities, dependencies);
- **Document**.

Hyperedges:

- **Local** (between each concept and the document).
- **Global** (between all concepts and a document).

- Ranking based on the **factor graph** representation of the hypergraph.
- Score computed through:
 - The combination of **log-linear models** for local and global factors;
 - Based on a **language modeling estimate** for each concept.



TAG: Text As Graph

- Proposed by Dekker and Birnbaum in 2017 as a hypergraph-based representation model for text.
- Built to mitigate XML limitations (e.g., lack of overlap).

Nodes:

- Document;
- Text;
- Markup;
- Annotation.

Hyperedges:

- Document-to-text;
 - Marks the beginning of the document.
- Text-to-text;
 - Links sequences of tokens.
- Annotation-to-text/-markup /-annotation;
 - Links text blocks to parts of the document.
- Markup-to-text/-markup;
 - Assigns types (possibly hierarchical).

**We also proposed a
hypergraph-based model for
the joint representation of text
and entities, and generalized
entity-oriented search.**





What is entity-oriented search?

“Entity-oriented search is the search paradigm of organizing and accessing information centered around entities, and their attributes and relationships.”

– Balog, Entity-Oriented Search, 2018.


Like Google!

 sci-fi movies from 1985 


[All](#) [Images](#) [Videos](#) [Shopping](#) [News](#) [More](#) [Settings](#) [Tools](#)

Movies > Sci-fi


All Genres < 1985 >




Back to the Future
1985




Lifeforce
1985




The Quiet Earth
1985




Mad Max Beyond Thunderdome
1985




D.A.R.Y.L.
1985




Enemy Mine
1985




Creature
1985




Explorers
1985




My Science Project
1985




Cocoon
1985




The Stuff
1985



Brazil
1985



Weird Science
1985



Galaxy Invaders
1985

1985 Sci-Fi Movies - MovieWeb

<https://movieweb.com/movies/1985/sci-fi/>

1985 Sci-Fi movies, movie release dates. A complete list of Sci-Fi movies in 1985.

Top 10 Sci Fi Movies of 1985 - IMDb

<https://www.imdb.com/list/ls056100695/>

Aug 5, 2013 - Brazil (1985) R | 132 min | Drama, Sci-Fi. Rate. Rate. Back to the Future (1985) PG | 116 min | Adventure, Comedy, Sci-Fi. 8.5. Re-Animator (1985) Unrated | 105 min | Comedy, Horror, Sci-Fi. 7.2. Real Genius (1985) PG | 108 min | Comedy, Romance, Sci-Fi. Rate. Enemy Mine (1985) PG-13 | 108 min | Action, Adventure,

Or Bing!


Metallica


<http://metallica.com>


OVERVIEW


TOP SONGS


ALBUMS


 Wikipedia

 Instagram


 Twitter

 Facebook

 YouTube


 Myspace

News about Metallica




Metallica Guitarist Kirk Hammett: How to Stay in the Moment During Big Performances

Kirk Hammett, lead guitarist for Metallica and co-founder of KHDK Electronics, describes to Inc. contributing ...
Inc. · 1d




Metallica, Fantasia talk anthems at NBA Finals

OAKLAND, Calif. — Kirk Hammett of Metallica says with all he's accomplished in his career, playing "The Star-Spangled Banner" at the ...
YAHOO! · 13h



Metallica join locals in Clare pub and take in a stroll along the Cliffs of Moher

Extra.ie on MSN.com · 1d




Metallica, Fantasia talk anthems at NBA Finals

OAKLAND, Calif. — Kirk Hammett of Metallica says with all he's accomplished in his career, playing "The Star-Spangled Banner" at the ...
The Washington Post · 5d


[See more news about Metallica](#)

Videos of metallica




Metallica - Enter Sandman [Official Music Video]

333M views
YouTube · 3/19/2012



Metallica - One [Official Music Video]

128M views
YouTube · 10/27/2009




Metallica - Nothing Else Matters [Official Music Video]

616M views
YouTube · 10/27/2009


[See more videos of metallica](#)

Videos of metallica




Metallica - Enter Sandman [Official Music Video]

333M views
YouTube · 3/19/2012



Metallica - One [Official Music Video]

128M views
YouTube · 10/27/2009



Metallica - Nothing Else Matters [Official Music Video]


616M views
YouTube · 10/27/2009

[See more videos of metallica](#)

Top songs


The Unforgiven

Metallica · 1991



Enter Sandman

Metallica · 1991




One

...and Justice for All · 1988


Nothing Else Matters

Metallica · 1991



Master of Puppets


1986




Whiskey in the Jar

[See more tracks](#)


Albums




Helping Hands...Live...
2019




Live in the 90s
2018




Metallica
1991



Hardwired... to Self-Deception
2016




Master of Puppets
1986







Kill 'Em All
1983

[See more albums](#)



Metallica | Twitter

[@Metallica](#) ·  

TONIGHT IN COLOGNE  

[talli.ca/metincologne-4](#) Ghost and Bokassa kick off the show! Poster Art: Ken Taylor Enhanced Experience Poster Art: Paige Reynolds

8 hours ago

Cologne! Coming down to the show tonight at RheinEnergieStadion? Use the hashtag #MetinCologne and your posts may be featured on Metallica.com! ... [#WorldWired](#)

9 hours ago


Miss out on a poster at yesterday's show in Amsterdam? A very limited number of Acom's prints are available online now. [Metallica.com](#) ➡ [talli.ca/2WCjhjm](#) Probitly Merch ➡ [talli.ca/2VH2lRo](#)

21 hours ago

Or even ANT!
(Shameless advertising)



<http://ant.fe.up.pt>




Todos Salas Pessoal Estudantes Notícias Ferramentas de Pesquisa

4 resultados (3.81 segundos)

José Luís da Silva Devezas

Sala
I123



Estudante

https://sigarra.up.pt/flup/pt/vld_entidades_geral.entidade_pa...
Faculdade de Belas Artes da Universidade do Porto (FBAUP) (mais 4)
Curso: Doutoramento em Informática
Código: 200303288

José Luís da Silva Devezas

Pessoal

https://sigarra.up.pt/feup/pt/func_geral.FormView?P_CODIGO...
Faculdade de Engenharia da Universidade do Porto (FEUP) (mais 1)
Código: 493720 **Sigla:** JLSD
1317

I123

Sala

https://sigarra.up.pt/feup/pt/instal_geral.espaco_view?pv_id=...
Laboratório de I&D de Sistemas de Informação, Faculdade de Engenharia da Universidade do Porto (FEUP)
Responsáveis: Carla Alexandra Teixeira Lopes
Edifício: Electrotecnia (I) **Andar:** 1



The tasks

Entity-oriented search tasks

Ad hoc document retrieval (using entities)

Input

Keyword Query: **croft** **bendersky**

Document mentioning the **College of Information and Computer Sciences** and **Hypergraph** entities, since **W. Bruce Croft** is dean of the **College of Information and Computer Sciences** and **Hypergraph** is one of the topics covered by **Michael Bendersky** in his thesis.

Output

Doc:
342

Doc:
13

Doc:
671

Faculty member and former Dean in the **College of Information and Computer Sciences.**

Recent Ph.D. Graduates:

Michael Bendersky
Van Dang



Entity-oriented search tasks

Ad hoc entity retrieval

Input

Keyword Query: croft bendersky

Output

Entity: [Person] W. Bruce Croft

Entity: [Person] Michael Bendersky



Entity-oriented search tasks

Related entity finding

Input

Entity: [Person] Michael Bendersky

Type: [ScholarlyArticle]

Relation: [creator]

Output

Entity: [ScholarlyArticle] Discovering key concepts in verbose queries

Entity: [ScholarlyArticle] Modeling higher-order term dependencies in information retrieval using query hypergraphs



Entity-oriented search tasks

Entity list completion

Input

Entity: [Person] Michael Bendersky

Type: [ScholarlyArticle]

Relation: [creator]

Example 1: [ScholarlyArticle] Information retrieval with query hypergraphs

Output

Entity: [ScholarlyArticle] Modeling higher-order term dependencies in information retrieval using query hypergraphs

Entity: [ScholarlyArticle] Discovering key concepts in verbose queries

This is more similar to the example, so we moved it up.

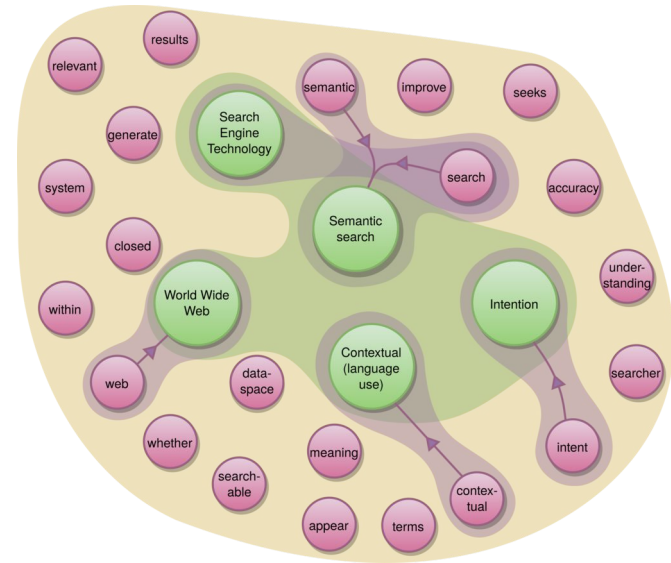
A general solution

Devezas, J., & Nunes, S. (2019). Hypergraph-of-entity: A unified representation model for the retrieval of text and knowledge. Open Computer Science Journal. *[Accepted to be published]*

Hypergraph-of-entity

Representation model

- Collection-based hypergraph.
- Joint representation of terms, entities and their relations.
- For indexing combined data (e.g., corpora linked to knowledge bases).



Devezas, J., & Nunes, S. (2019). Hypergraph-of-entity: A unified representation model for the retrieval of text and knowledge. Open Computer Science Journal. *[Accepted to be published]*

Hypergraph-of-entity

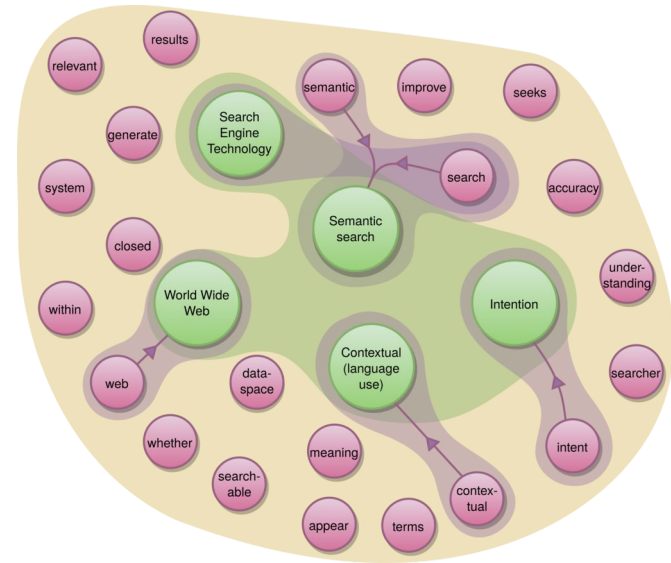
Representation model

Nodes:

- Terms;
- Entities.

Hyperedges:

- Document;
- Related to;
- Contained in.



Devezas, J., & Nunes, S. (2019). Hypergraph-of-entity: A unified representation model for the retrieval of text and knowledge. Open Computer Science Journal. *[Accepted to be published]*

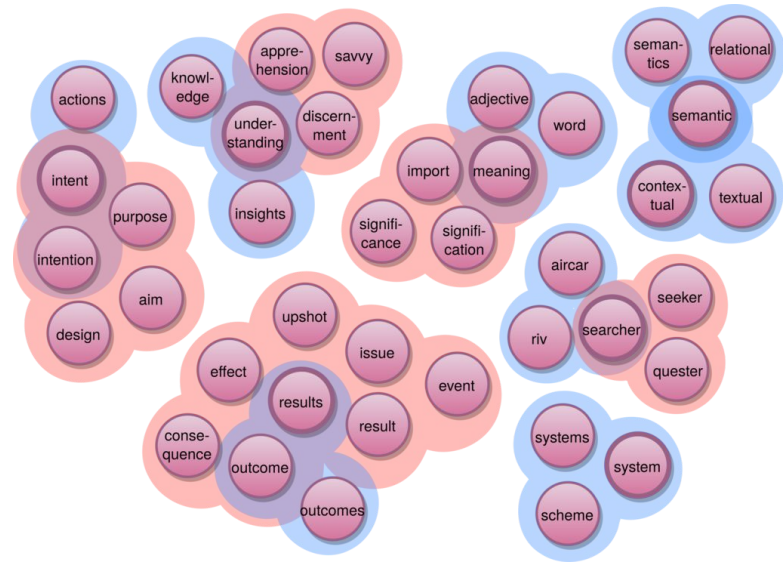
Hypergraph-of-entity

Representation model

We can easily add hyperedges for:

- Synonyms;
 - Based on WordNet SynSets.
- Contextual similarity.
 - Based on word2vec similarities.

Note: the remaining hyperedges and nodes are not displayed here to improve legibility.

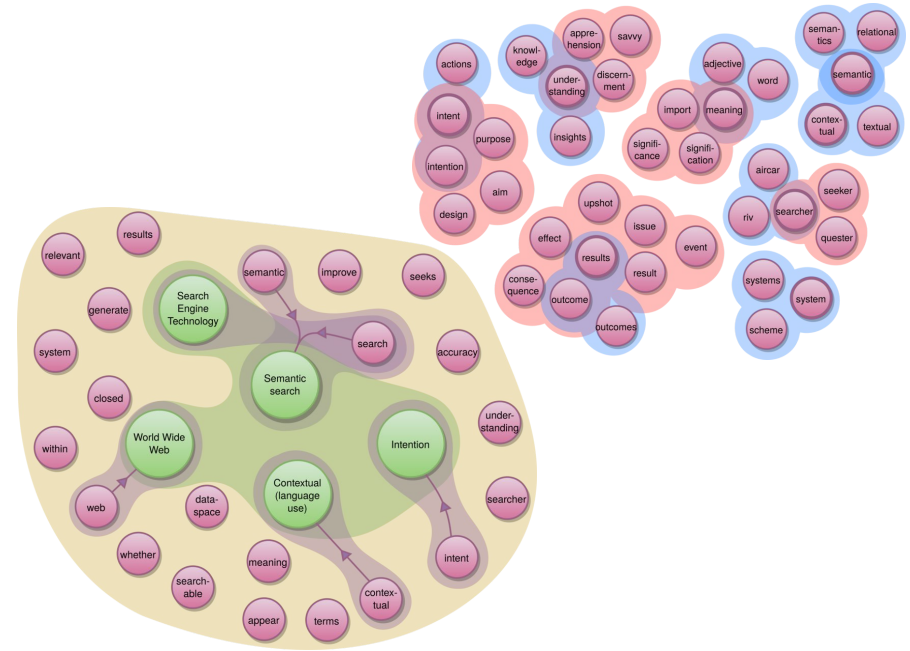


Hypergraph-of-entity

Ranking model

- Random walks departing from seeds assign a visitation frequency to nodes and hyperedges.
- Depending on the task, input and output changes, but the ranking function remains unchanged.
- Nodes or hyperedges are collected and ordered by visitation frequency to provide a ranking.

Devezas, J., & Nunes, S. (2019). Hypergraph-of-entity: A unified representation model for the retrieval of text and knowledge. Open Computer Science Journal. *[Accepted to be published]*



Wrapping up...



Final remarks

- Hypergraphs are expressive and flexible data structures, ideal for the joint representation of unstructured corpora and structured knowledge bases.
- They are also a way to harness higher-order relations in information retrieval.



Final remarks

- In the hypergraph-of-entity, a document hyperedge is like a bag of words and entities. Adding synonym or contextual similarity hyperedges creates new semantic bridges between documents.
- Random walks in graphs have historical relevance in web search engines (e.g., PageRank). Applying them to hypergraphs was a natural step.

Thank you!

You can experiment with the hypergraph-of-entity and other retrieval models using our evaluation framework, Army ANT:

<https://github.com/feup-infolab/army-ant>

Also available as a Docker image:

<https://github.com/feup-infolab/army-ant-install>
