BIBLME RecSys: Harnessing Bibliometric Measures for a Scholarly Paper Recommender System

Anaïs OLLAGNIER

Research Context

Viewed pages in 2014

76 322 739 France
35 124 966 United States
24 934 420 Deutschland
6 965 519 Bulgaria
1 548 991 Spain

Per million population

1 151 885
309 887
21 853
4 839
667
0

Source: OpenEdition
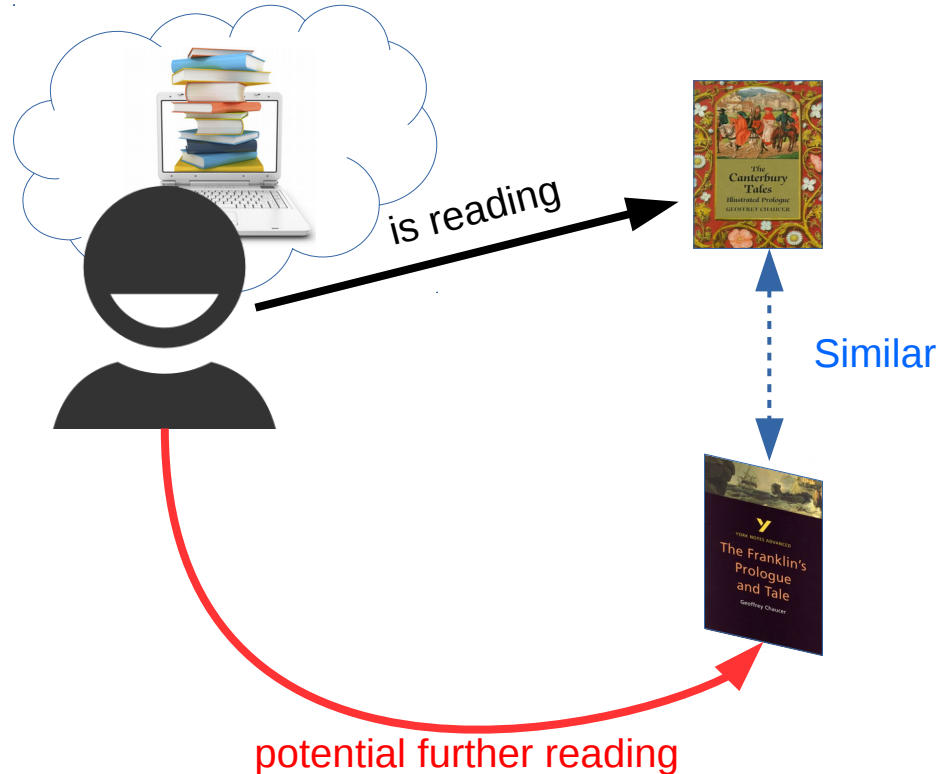©H.Théry OpenEdition 2015
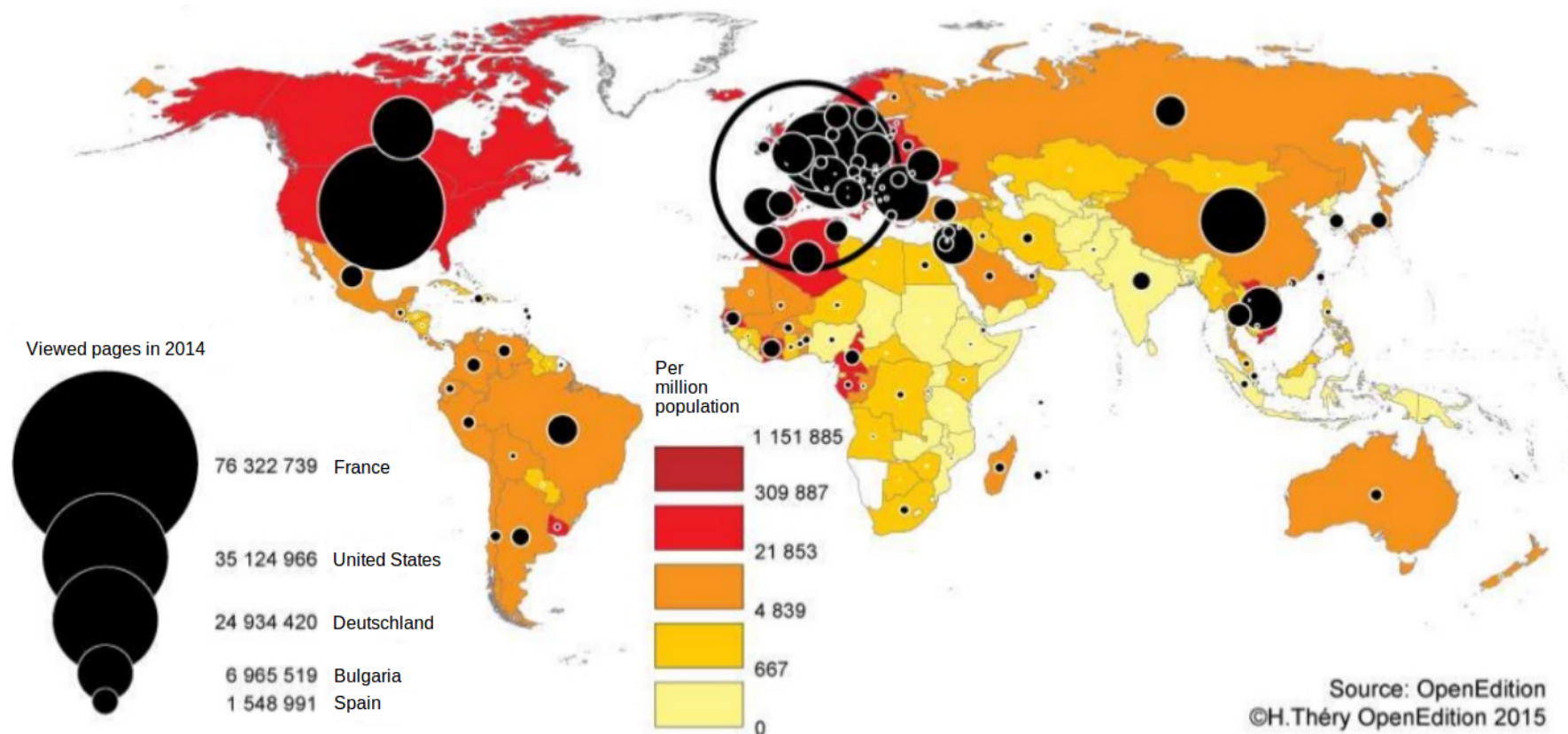
ANR program "Investissements d'Avenir" EquipEx DILOH (ANR-11-EQPX-0013)

Research Objectives

Develop a system Research Paper Recommender System in the context of a digital library.
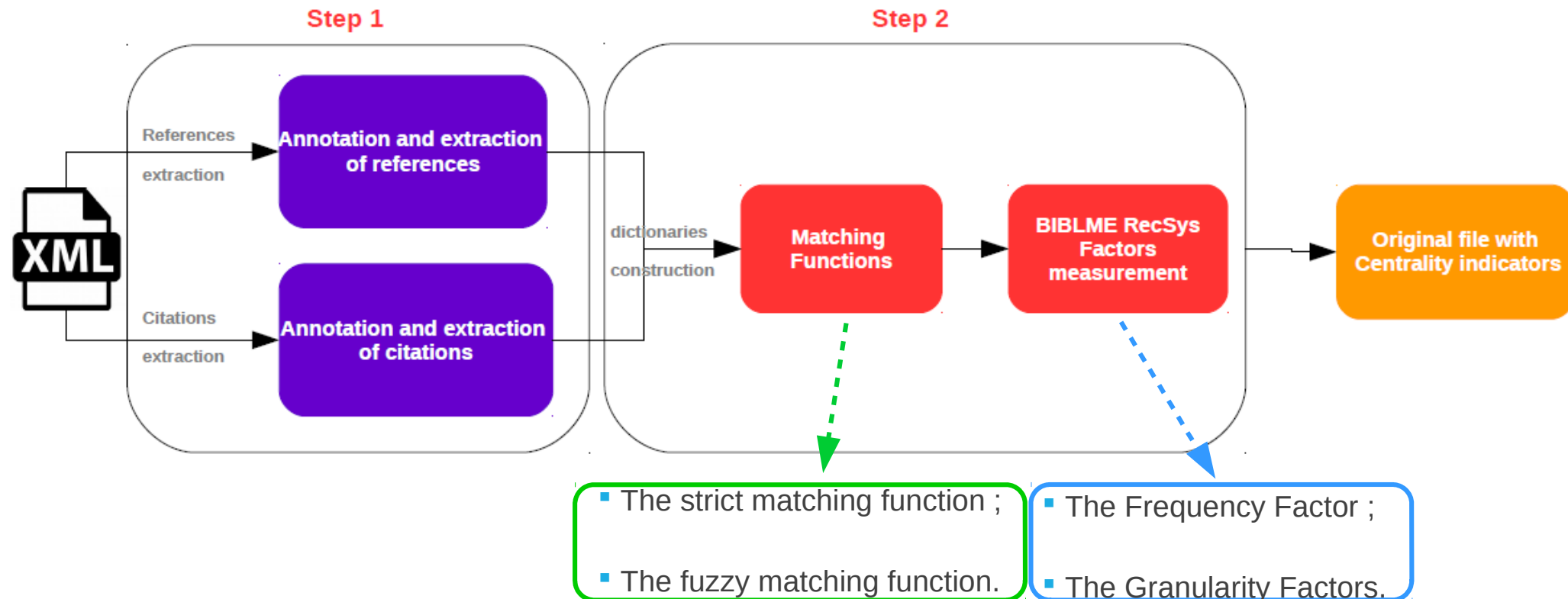
is reading

Similar

potential further reading

Specificities :

- guest
- full-text (XML TEI)
  - content
  - descriptive metadata
  - structural metadata

BiblMe RecSys : centrality indicator

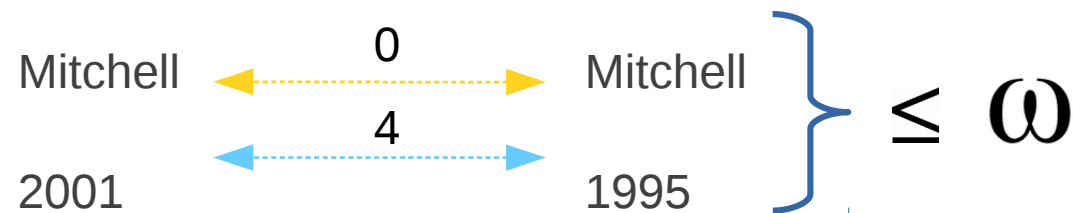Provide users with paper suggestions guided by the papers they are reading and by the references they contain.

- The strict matching function ;
- The fuzzy matching function.

- The Frequency Factor ;
- The Granularity Factors.

BiblMe RecSys : matching functions

- **The strict matching function** checks whether bibliographic fields of two references match.

(Mitchell, 1996, p.131)

(Mitchell, 1995, 2000, 2003)

Howard (1902)

(Howard, 1902, p.12)

- **The fuzzy matching function** corresponds to the Levenshtein distance.

Mitchell ←→ 0 ←→ Mitchell
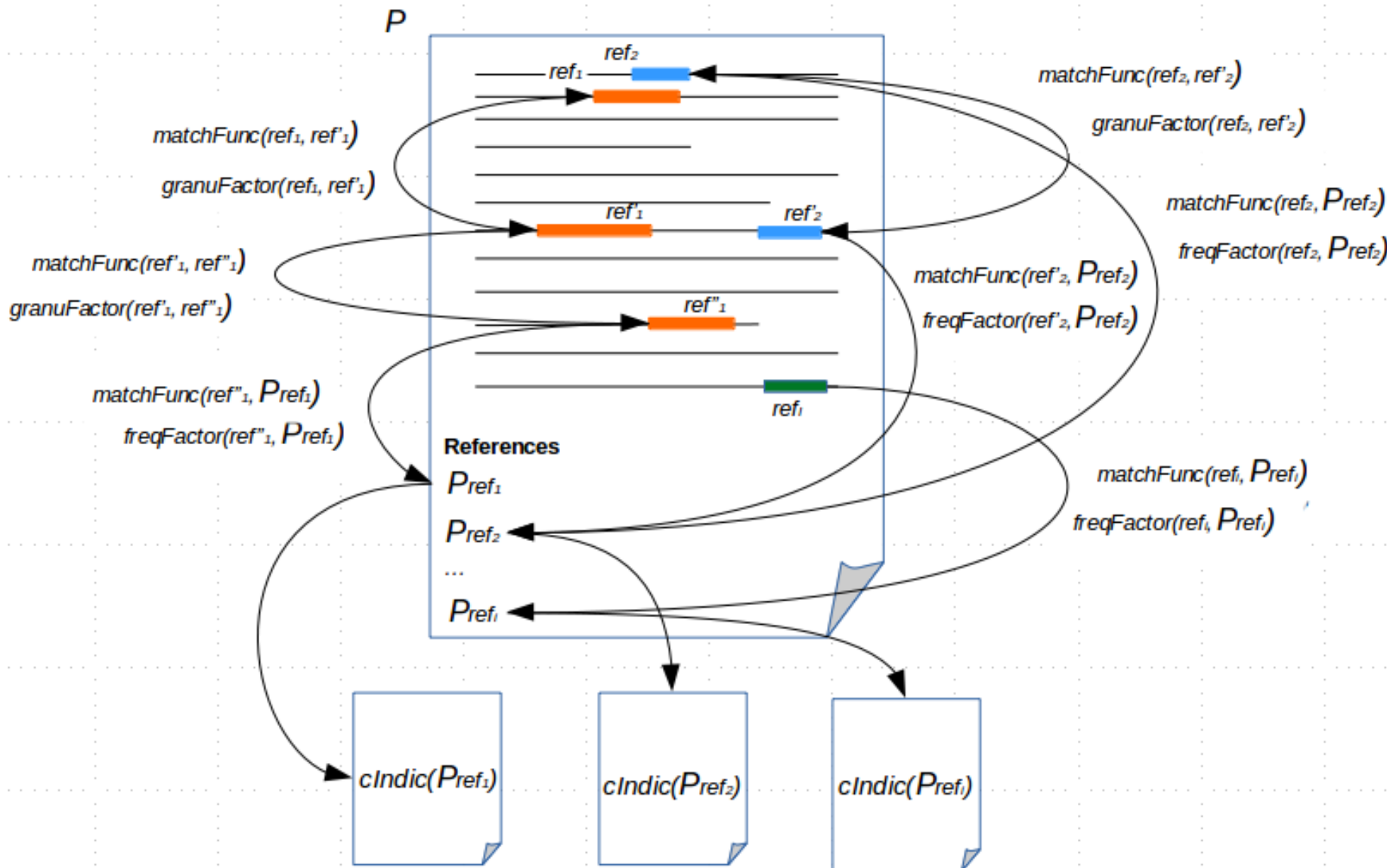
2001 ←→ 4 ←→ 1995

$$\leq \omega$$

BiblMe RecSys : centrality factors

> References' importance increase proportionally with numbers of mentions and more detailed discussion of the cited document (Tang, 2008).

- **The frequency factor** is based on citation counting.

- **The granularity factors** allow to distinguish how references are used by the authors :

  - The fine granularity: citations' concentration within low textual density areas tends to strengthen author's arguments on specic aspects of his researches.

  - The coarse granularity: more citations of a reference occur throughout the paper more the author is influenced by this work;

P

$matchFunc(ref_1, ref'_1)$

$granuFactor(ref_1, ref'_1)$

$matchFunc(ref'_1, ref''_1)$

$granuFactor(ref'_1, ref''_1)$

$matchFunc(ref''_1, P_{ref_1})$

$freqFactor(ref''_1, P_{ref_1})$

$matchFunc(ref_2, ref'_2)$

$granuFactor(ref_2, ref'_2)$

$matchFunc(ref_2, P_{ref_2})$

$freqFactor(ref_2, P_{ref_2})$

$matchFunc(ref'_2, P_{ref_2})$

$freqFactor(ref'_2, P_{ref_2})$

$matchFunc(ref_i, P_{ref_i})$

$freqFactor(ref_i, P_{ref_i})$

ref_2
ref_1
ref'_1
ref'_2
ref''_1
ref_i

References
P_{ref_1}
P_{ref_2}
...
P_{ref_i}

$cIndic(P_{ref_1})$
$cIndic(P_{ref_2})$
$cIndic(P_{ref_i})$

- Candidate paper to recommend (P_{ref_i})
- Centrality indicator ($cIndic(P_{ref_i})$)
- The matching functions ($matchFunc(ref_i, ref_j)$)
- The frenquency factor ($freqFactor(ref_i, P_{ref_i})$)
- The granularity factors ($granuFactor(ref_i, ref_j)$)

Experiment : Search OpenEdition

- Query (First author + Title)

> Alvarez The Relevance of Informational Infrastructures in Future Cities | All fields

SEARCH   RESET   CREATE A NOTIFICATION ASSOCIATED WITH THIS SEARCH

13 results on 1 page(s)   1

Belgeo

Inaugural issue
>> http://journals.openedition.org/belgeo/743

- Proposed results

... of the Belgian National Committee for Geography in association with the Société Royale Belge ... social elites in the 19th and the early 20th century. As a matter of fact, classical geography has been ... methods and skills enabled to take stock of the resources in both potential and acquired colonies. After ... merging with its academic counterpart, the Institute of British Geographers in 1995. Not all large ... debate in Geoforum (2005). A completely different attempt towards integration, is the introduction of ... in their third and last edition of the International List of Geographical Serials, a number that ... Cities and the Limitations of Geographical Scale Analysis", in SHEPPARD E. and MASTER R. B. (Eds.) (2004 ... A. (2000), "Belgian geographers at work: the occupations of graduates in 1999", BELGEO - Belgian ... . (2004), "Geography in Belgium" , Belgeo, 5, 1, pp. 33-44. PITMAN A.J. (2005), "On the role of Geography ...

Publication
Belgeo
Publication type: ̦ Revues • Document type: ̦ Numéro de revue
Authors
Manuel B. Aalbers, Carles Carreras I Verdaguer, Hugh Clout, Camille Cotteels, Frédéric Dobruszkes, Yann Dubois, Aurora Garcia Ballesteros, Maria Dolors Garcia Ramon, Claude Grasland, Jean-Marie Halleux, Jonathan Jones, Christian Kesteloot, Walter Matznetter, Armando Montanari, Robert Musil, Anna Ortiz, Dominique Peeters, Herminia Pujol, Patrick Rérat, Ugo Rossi, Serge Schmitz, Isabelle Thomas, Christian Vandermotten, Thomas Vanoutrive, Ties Vanthillo, Herman van der Wusten, Ann Verhetsel, Edwin Zaccai
Publication date
mars 2012
Document availability
Abstract available

- OpenEdition search engine:

  Apache Lucene retrieval model

  Boolean filter

  BM25 similarity

Experiments : Evaluation platform

- Two lists of five papers are proposed for each system ;

  - Selection of the most relevant proposal ;
  - Refining evaluation through an assessment ;
  - Possibility to make comments.

Rank the selection

You have selected "Gothique, réforme et Panoptique. Wrobel"

Choose the proposal that is most relevante as a whole

Proposition 1 ▼

Proposition 1:

		Proposition 2:	
Surveiller et punir Foucault	0 ▼	Gothique, réforme et Panoptique	0 ▼
L'œil du pouvoir Foucault	0 ▼	Varia	0 ▼
The Castle of Otranto Walpone	0 ▼	Varia	0 ▼
The Old English Baron Reeve	0 ▼		
Nights at the Circus Carter	0 ▼		

Add a suggestion

Submit

Experiments : Results

Paper	BiblMe RecSys recommendation	SoE recommendation
Jaubert - Correspondance as an Ethical Genre	7	5
Bisson - Sufism and Tradition	8	0
Nabti - Sufis in Parisian suburbs	5	0
Laborde - The hermit and the virtuoso	6	0
Esquenazi - From star system to people	4	4
Danou - On a Novella of Arthur Schnitzler (1862-1931), Flight into Darkness	2	1
Wrobel - Gothic, Reform and Panoptic	3	0
Amiraly - The impact of a pilot water metering project in an Indian city on users perception of the public water supply	4	0
Masdonati - The question of identity in the dual Swiss vocational training system: a contrasting picture	1	5
Delannoy - Karst : from palaeogeographic archives to environmental indicators	1	0
Duval - The deceased to the shackles	2	2
Angevin - The Magdalenian lithic industry from the openair site of la Corne-de-Rollay (Couleuvre, Allier): production standards andproduction lines variability	2	0

Conclusions and perspectives

Conclusions :

- **Paper suggestions** guided by the paper the user is reading and by the references it contain ;
- Predict **references' importance** through textual features ;
- Recommend papers according to their **centrality indicator** .

Perspectives :

- Examine the degree of users' satisfaction ;
- Exploit recommendation algorithms by graph analysis ;
- Determine relations between references and their proximity .

Thanks for your attention

BiblMe RecSys

In short digital technologies, which McLuhan described as an extension of our nervous system (McLuhan, 1964), will be integrated into large "internet of things" architectures [...] "digital nervous systems" of urban, even planetary scales (Mitchell, 1995, 2000, 2003). […] "upgrading" of traditional cities infrastructures has created a '$100 billion jackpot' (Townsend, 2013, pp. 19), one that has fueled an industrial race for the transformation of the next generation of urban infrastructures. […] in the following decades of explosive urban growth and reconfiguration (Townsend, 2013). […] through a variety of digital augmentations (Wiig, 2015), and question the self-congratulatory rhetoric surrounding them, their fuzzy definition and overall ideology (Holland, 2005). […] "As traditional cities have evolved so have customs, norms and laws governing rights to privacy, access to public and semi public places and exertion of control" (Mitchell, 1996, p.131)

- Mitchell, W. J. (2000). E-topia:" urban life, Jim--but not as we know it". MIT press.

- Townsend, A. M. (2013). Smart cities: Big data, civic hackers, and the quest for a new utopia. WW Norton & Company.

- Holland, R. (2005). Will the real smart city stand up. City, 12(3), 3.

- McLuhan, M. (1964).Understanding media: The extensions of man. MIT press.