COST Action IC1302

semantic KEYword-based Search on sTructured data sOurcEs

keystone

SCHOLARLY BOOK RECOMMENDATION : A TASK COMBINING SENTIMENT ANALYSIS, INFORMATION EXTRACTION AND IR

Patrice Bellot (M. Dacos, C. Benkoussas, A. Ollagnier, M. Orban, H. Hamdan, A. Htait, E. Faath, Y.M. Kim, F. Béchet)

Aix-Marseille Université - CNRS (LSIS UMR 7296 ; OpenEdition)

patrice.bellot@univ-amu.fr

LSIS - DIMAG team http://www.lsis.org/dimag

OpenEdition Lab : http://lab.hypotheses.org

OpenEdition

Aix*Marseille université

OpenEdition

1805 BOOKS | 393 JOURNALS | 1629 BLOGS | 28138 EVENTS

Journals catalogue | SEARCH

OpenEdition brings together four platforms dedicated to electronic resources and academic information in the humanities and social sciences. Your access to these services is provided by: Aix Marseille Université. By supporting OpenEdition, your institution is actively promoting Open Access scientific publishing.

Aix*Marseille université

FURTHER INFORMATION

OpenEdition books

Focus



PRINCIPES DE GÉOGRAPHIE HUMAINE

Publiés d'après les manuscrits de l'auteur par Emmanuel de





Our partners: libraries an institutions all over the world

Latest selections

18 SEPTEMBER 2014
WOMEN, GENDER AND SEXUALITY

06 MAY 2014
CULTURAL METISSAGE IN LATIN AMERICA

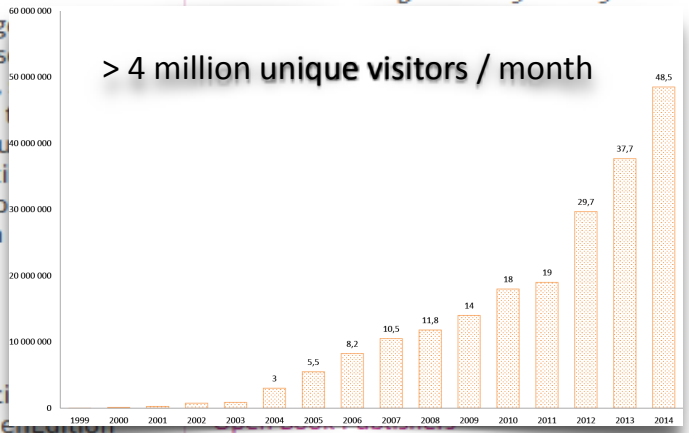
South America is heterog
heterogeneity is, of cours
to the vast geographical,
cultural area covered by
But heterogeneity also c
the South American conti
cause of the numerous fo
bridity, cross-fertilisation
that exis...

28 OCTOBER 2013
SYRIA

A major political and soci
is happening in Syria. Ope
Books invites you to explore the dif-
aussi transversale à l'aire sudaméri-

Latest excerpts

It now seems obvious, for example, that simple generalizations about the collective nature of art in non-literate cultures cannot hold good in face of the



Open Book Publishers

Some Open Questions

— **OpenEdition Lab** : a DH Research Program that aims to :
  — Detect hot topics, hot books, hot papers
  — Develop approaches for recommending and searching for books

— How to combine Information Retrieval and Recommendation (Learning to Rank ?)

— Is combining official meta-data and Reader-generated content (reviews) useful ?

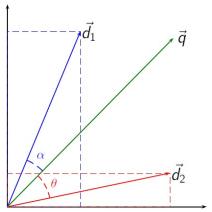
— How to identify and exploit bibliographic references (cross-linking) ?

— Is the content of the books useful ? (vs. reviews / abstracts only) ?

— How to deal with complex queries ?

— How to go beyond topical relevance ? (genre, reading level, recentness…)

Classical solution (the Vector Space Model for IR): a weight is associated to every word in every document (a vector), the query is represented as a vector in the document space, a degree of similarity between the documents and the query can be estimated



$$s(\vec{d},\vec{q}) = \sum_{i=1}^{i=n} w_{m_i,d} \cdot w_{m_i,q}$$

Searching for Books ?

Very diverse needs :

— **Topicality**

— With a precise context eg. *arts **in** China **during** the XXth century*

— With named entities : **locations** (the book is about a specific location OR the action takes place at this location), **proper names**…

— **Style / Expertise / Language**

— fiction, novel, essay, proceedings, position papers…

— for experts / for dummies / for children …

— in English, in French, in old French, in (very) local languages …

— **looking for citations** / **references**

— in what book appears a given citation

— what are the books that refer to a given one

— **Authority** :

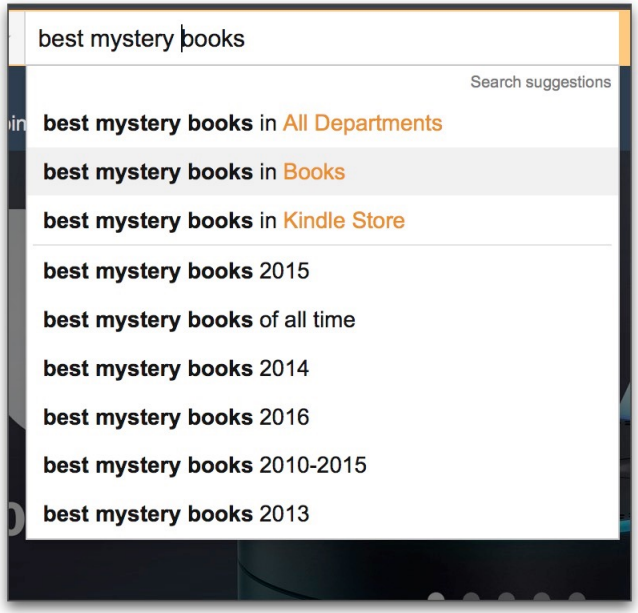
— What are the **most important** books about … (what *most important* means ?)

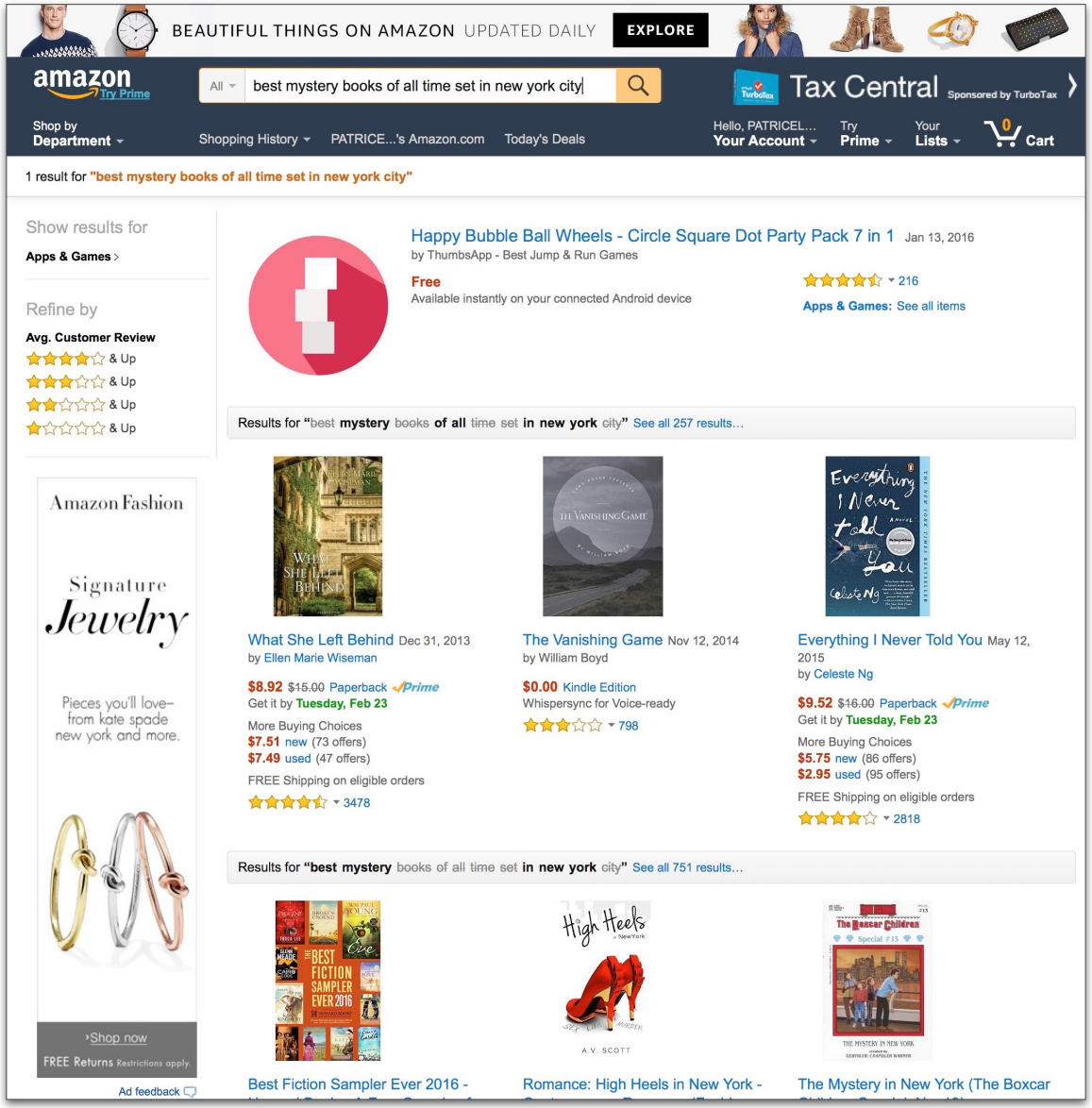
— What are the **most popular** books about …

For Digital Humanities (and for Research Purposes) : ***Researchers may not be interested in the best-sellers but in the impact of a book = in the debate it provokes***

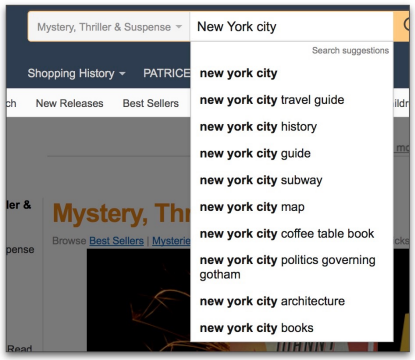
Searching for books : a complex (?) query

I am looking for the best mystery books set in New York City





Category : Mystery
Sort by Avg. Cust. Review
Query : New York

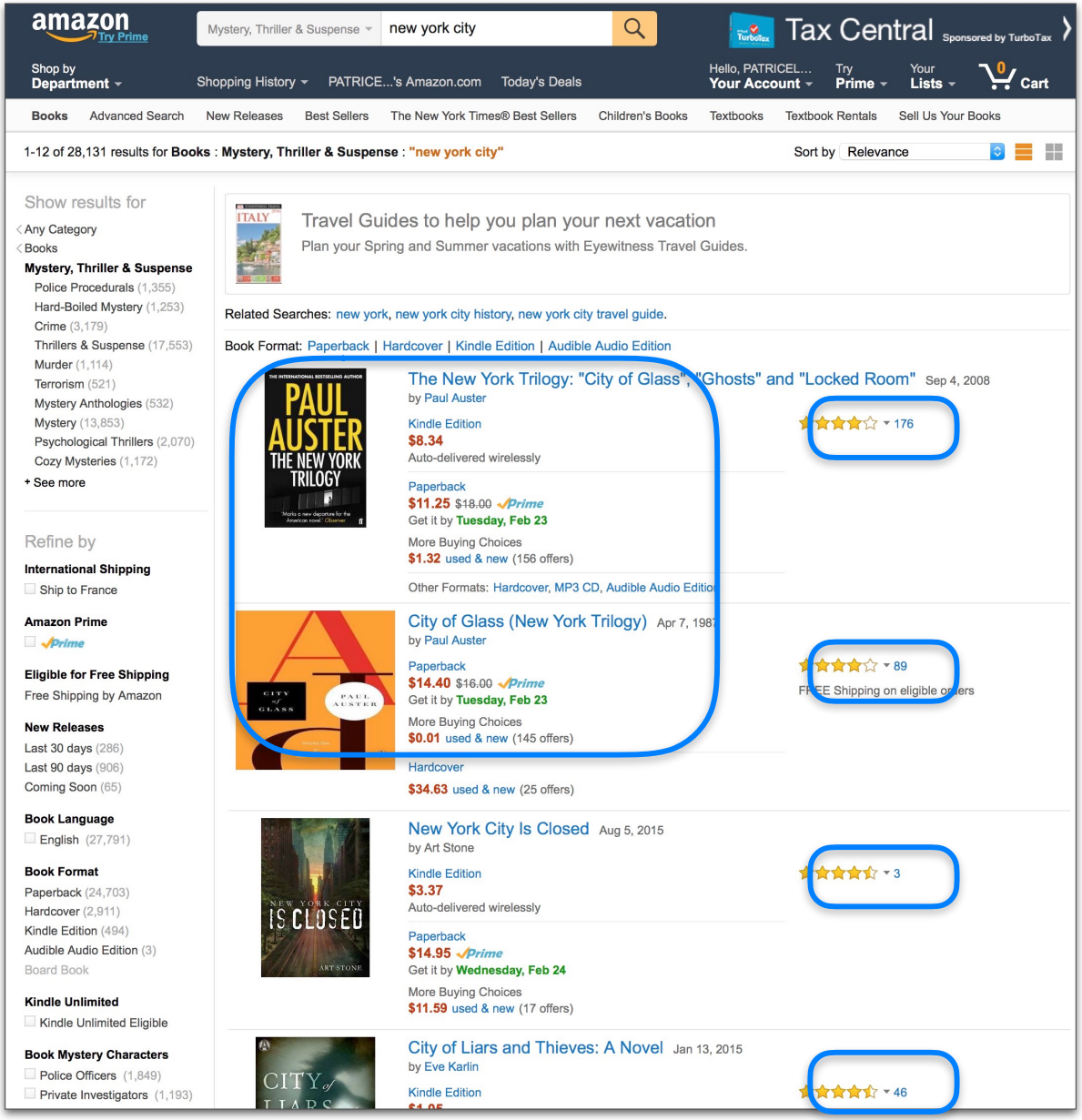


« New York city » occurs only once in the book. But published in NYC, reviewed by the NY Times

first position on The New York Times Best Seller list in 2013, « New York city » occurs only once in the book.

in Buffalo NY, near Toronto

Sort by *Relevance* ?



Social Book Search Lab

The Social Book Search Lab is part of CLEF 2016.

Organizers

Marijn Koolen (University of Amsterdam)
Toine Bogers (Aalborg University Copenhagen)
Antal van den Bosch (Radboud University Nijmegen)
Antoine Doucet (University of Caen)
Maria Gaede (Humboldt University Berlin)
Preben Hansen (Stockholm University)
Mark Hall (Edge Hill University)
Iris Hendrickx (Radboud University Nijmegen)
Hugo Huurdeman (University of Amsterdam)
Jaap Kamps (University of Amsterdam)
Vivien Petras (Humboldt University Berlin)
Michael Preminger (Oslo and Akershus University College of Applied Sciences)
Mette Skov (Aalborg University Copenhagen)
Suzan Verberne (Radboud University Nijmegen)
David Walsh (Edge Hill University)

News

31-01-2015: Suggestion Track topic set released!.

26-01-2015: Mining Track information and schedule updated!.

26-01-2015: Mining Track training data released!.

Table 1. Some facts about the Amazon collection.

Number of pages (i.e. books)	2,781,400
Number of reviews	15,785,133
Number of pages that contain a least a review	1,915,336

SBS Lab Overview

The Social Book Search (SBS) Lab investigates book search in scenarios where users search with more than just a query, and look for more than objective metadata. Real-world information needs are generally complex, yet almost all research focuses instead on either relatively simple search based on queries or recommendation based on profiles. The goal is to research and develop techniques to support users in complex book search tasks.

The Social Book Search Lab consists of three tracks:

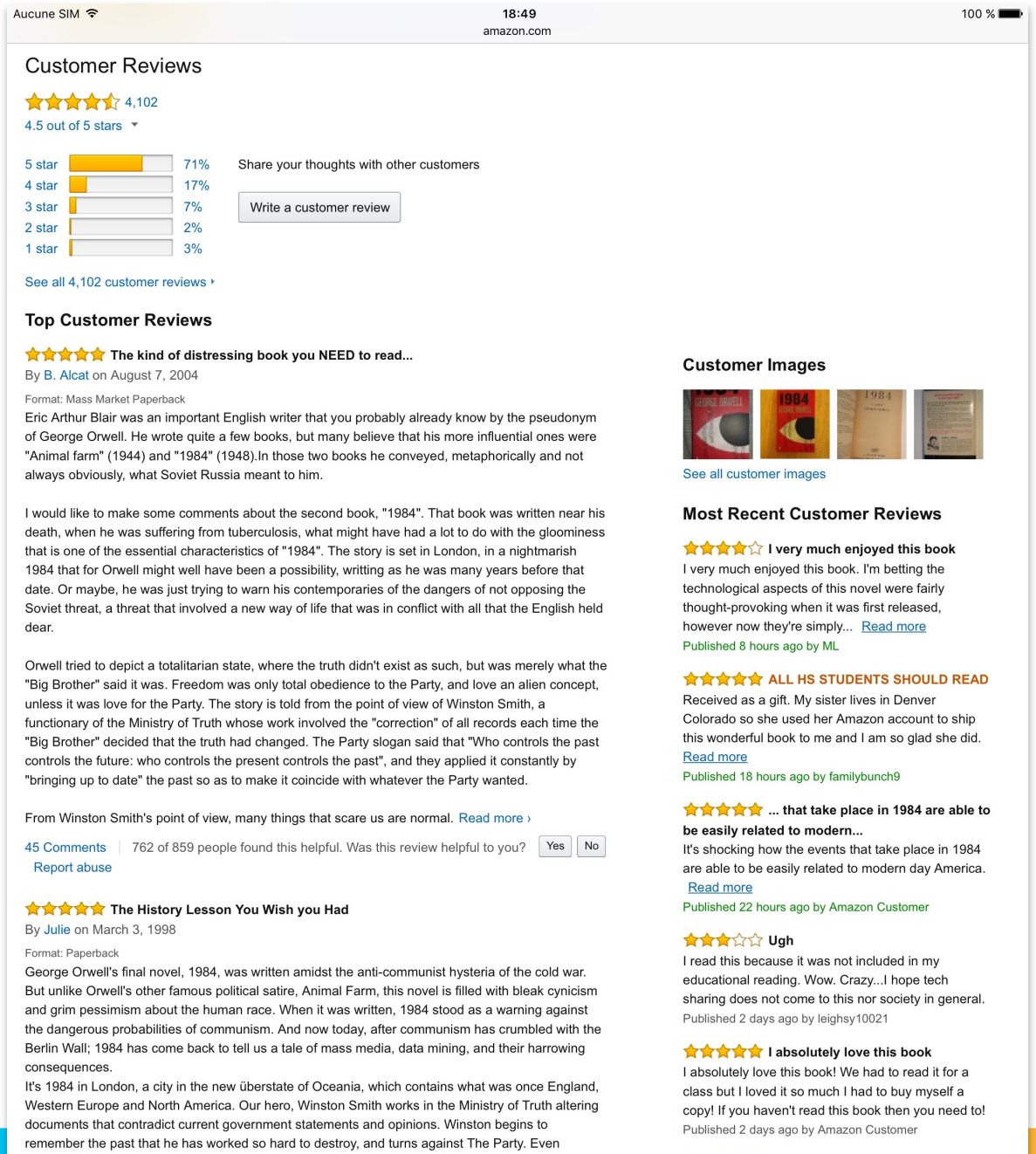
- Interactive Track: a user-oriented interactive task investigating systems that support users in each of multiple stages of a complex search tasks. The track offers participants a complete experimental interactive IR setup and an exciting new multistage search interface to investigate how users move through search stages.
- Suggestion Track: a system-oriented task for systems to suggest books based on rich search requests combining several topical and contextual relevance signals, as well as user profiles and real-world relevance judgements.
- Mining Track: an NLP/Text Mining track focussing on detecting and linking book titles in online book discussion forums, as well as detecting book search research in forum posts for automatic book recommendation.

If you're interested in participating in this task, please register as a participant at the CLEF 2016 Labs.

If you want to be informed of news and discussions, join our mailinglist.

http://social-book-search.humanities.uva.nl/#/overview

Searching for Books : Looking at Reviews



Aucune SIM 18:49 amazon.com 100 %

Customer Reviews

4,102

4.5 out of 5 stars

5 star	71%
4 star	17%
3 star	7%
2 star	2%
1 star	3%

Share your thoughts with other customers

Write a customer review

See all 4,102 customer reviews

Top Customer Reviews

The kind of distressing book you NEED to read...

By B. Alcat on August 7, 2004

Format: Mass Market Paperback

Eric Arthur Blair was an important English writer that you probably already know by the pseudonym of George Orwell. He wrote quite a few books, but many believe that his more influential ones were "Animal farm" (1944) and "1984" (1948).In those two books he conveyed, metaphorically and not always obviously, what Soviet Russia meant to him.

I would like to make some comments about the second book, "1984". That book was written near his death, when he was suffering from tuberculosis, what might have had a lot to do with the gloominess that is one of the essential characteristics of "1984". The story is set in London, in a nightmarish 1984 that for Orwell might well have been a possibility, writting as he was many years before that date. Or maybe, he was just trying to warn his contemporaries of the dangers of not opposing the Soviet threat, a threat that involved a new way of life that was in conflict with all that the English held dear.

Orwell tried to depict a totalitarian state, where the truth didn't exist as such, but was merely what the "Big Brother" said it was. Freedom was only total obedience to the Party, and love an alien concept, unless it was love for the Party. The story is told from the point of view of Winston Smith, a functionary of the Ministry of Truth whose work involved the "correction" of all records each time the "Big Brother" decided that the truth had changed. The Party slogan said that "Who controls the past controls the future: who controls the present controls the past", and they applied it constantly by "bringing up to date" the past so as to make it coincide with whatever the Party wanted.

From Winston Smith's point of view, many things that scare us are normal. Read more ›

45 Comments | 762 of 859 people found this helpful. Was this review helpful to you? Yes No

Report abuse

The History Lesson You Wish you Had

By Julie on March 3, 1998

Format: Paperback

George Orwell's final novel, 1984, was written amidst the anti-communist hysteria of the cold war. But unlike Orwell's other famous political satire, Animal Farm, this novel is filled with bleak cynicism and grim pessimism about the human race. When it was written, 1984 stood as a warning against the dangerous probabilities of communism. And now today, after communism has crumbled with the Berlin Wall; 1984 has come back to tell us a tale of mass media, data mining, and their harrowing consequences.
It's 1984 in London, a city in the new überstate of Oceania, which contains what was once England, Western Europe and North America. Our hero, Winston Smith works in the Ministry of Truth altering documents that contradict current government statements and opinions. Winston begins to remember the past that he has worked so hard to destroy, and turns against The Party. Even

Customer Images

See all customer images

Most Recent Customer Reviews

I very much enjoyed this book

I very much enjoyed this book. I'm betting the technological aspects of this novel were fairly thought-provoking when it was first released, however now they're simply... Read more

Published 8 hours ago by ML

ALL HS STUDENTS SHOULD READ

Received as a gift. My sister lives in Denver Colorado so she used her Amazon account to ship this wonderful book to me and I am so glad she did. Read more

Published 18 hours ago by familybunch9

... that take place in 1984 are able to be easily related to modern...

It's shocking how the events that take place in 1984 are able to be easily related to modern day America. Read more

Published 22 hours ago by Amazon Customer

Ugh

I read this because it was not included in my educational reading. Wow. Crazy...I hope tech sharing does not come to this nor society in general.

Published 2 days ago by leighsy10021

I absolutely love this book

I absolutely love this book! We had to read it for a class but I loved it so much I had to buy myself a copy! If you haven't read this book then you need to!

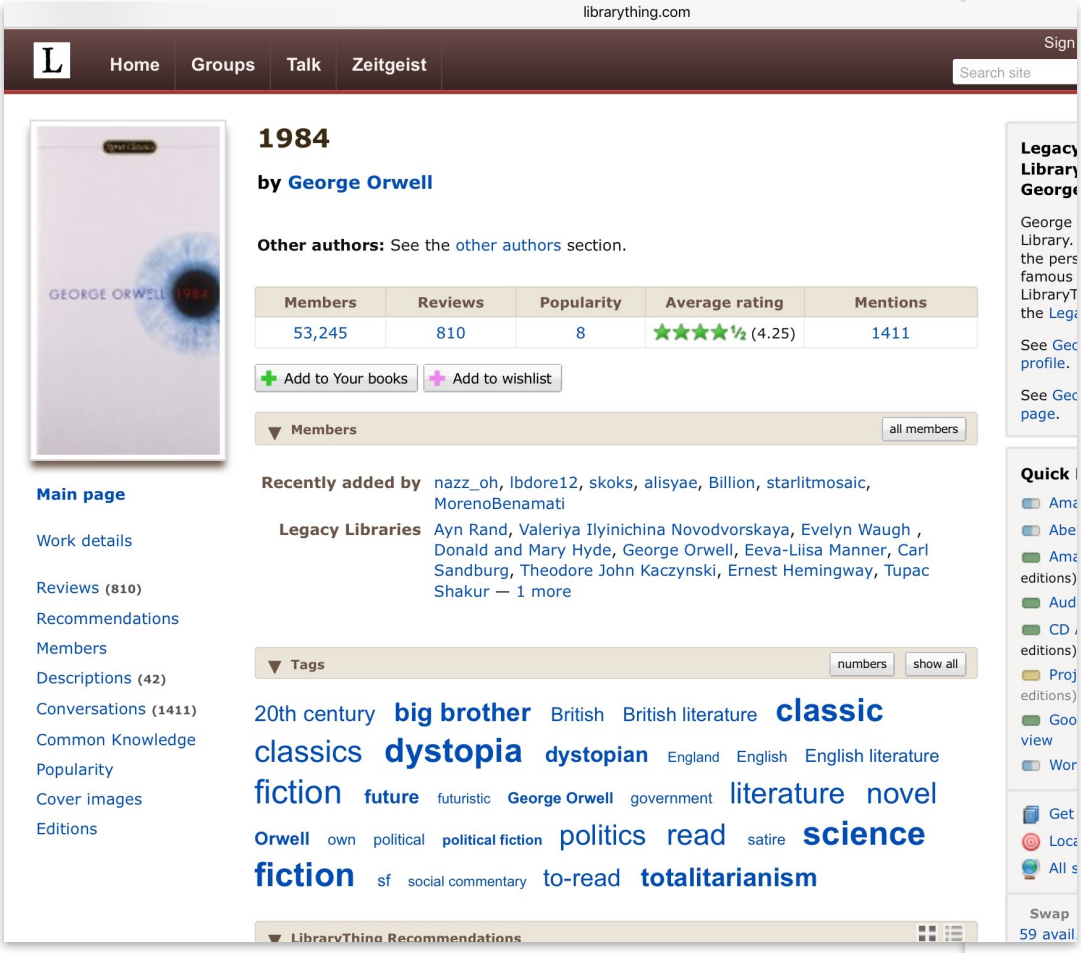
Published 2 days ago by Amazon Customer

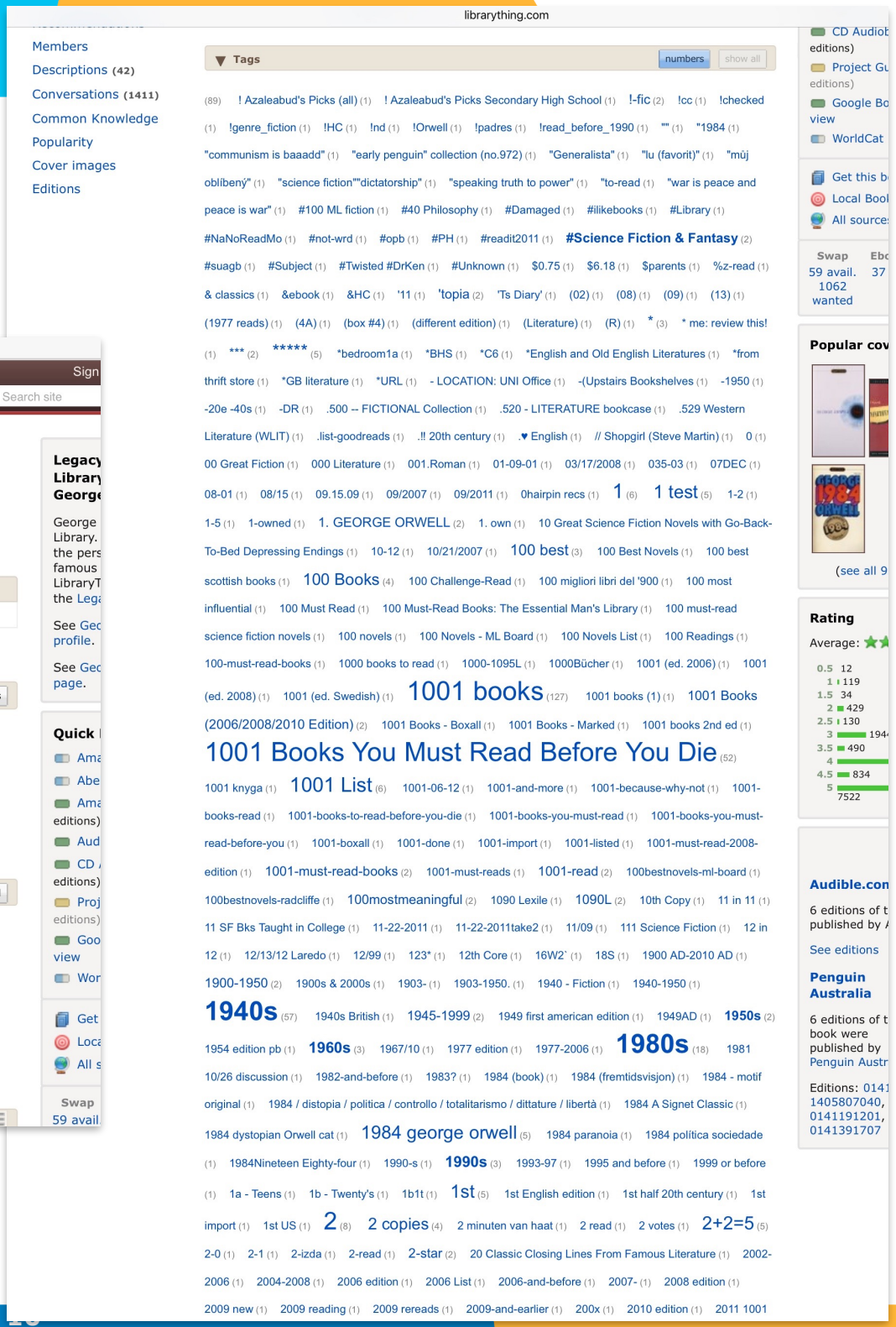
Reviews contain :

- keywords
- topics
- sentiment
- abstracts
- other books

Aix*Marseille université

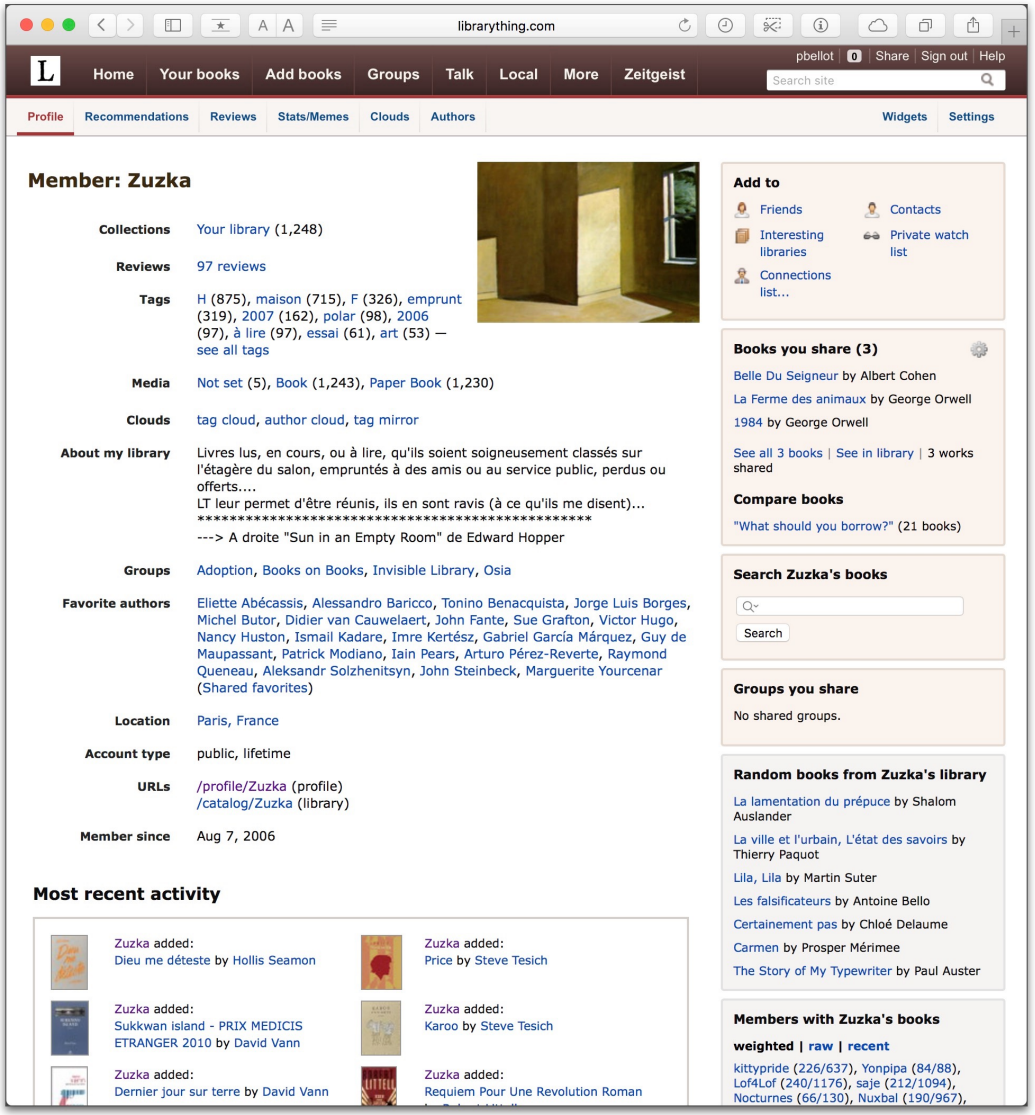
Social Tagging

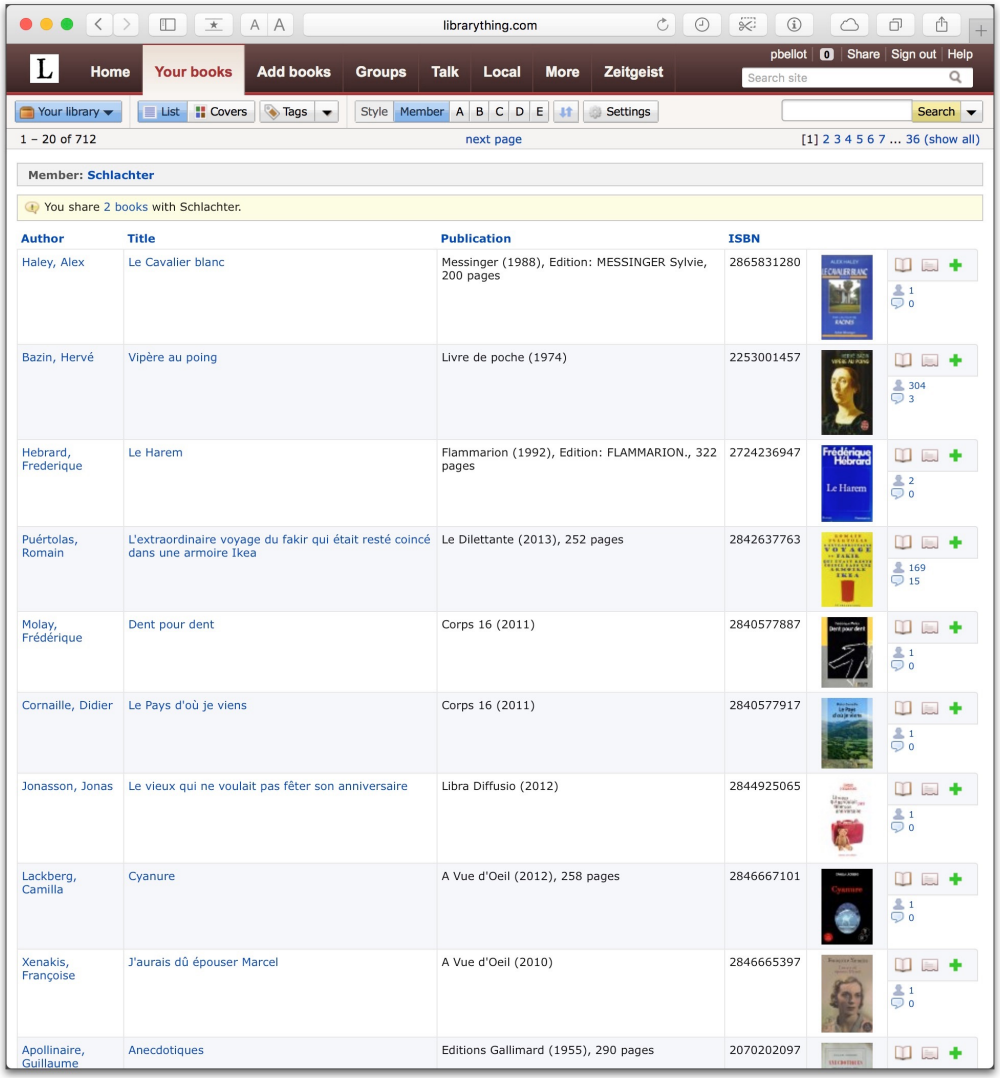




Complement categories but a lot of tags !

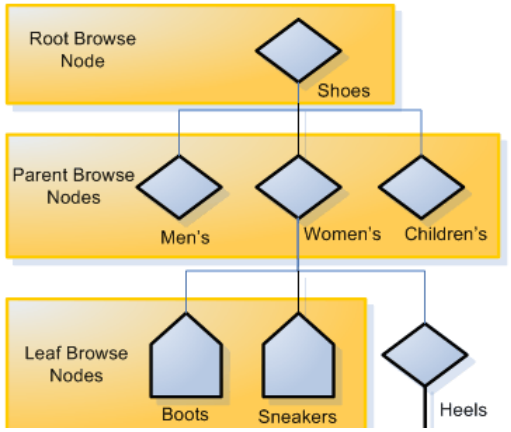
User Profiles (catalog, reviews, ratings)



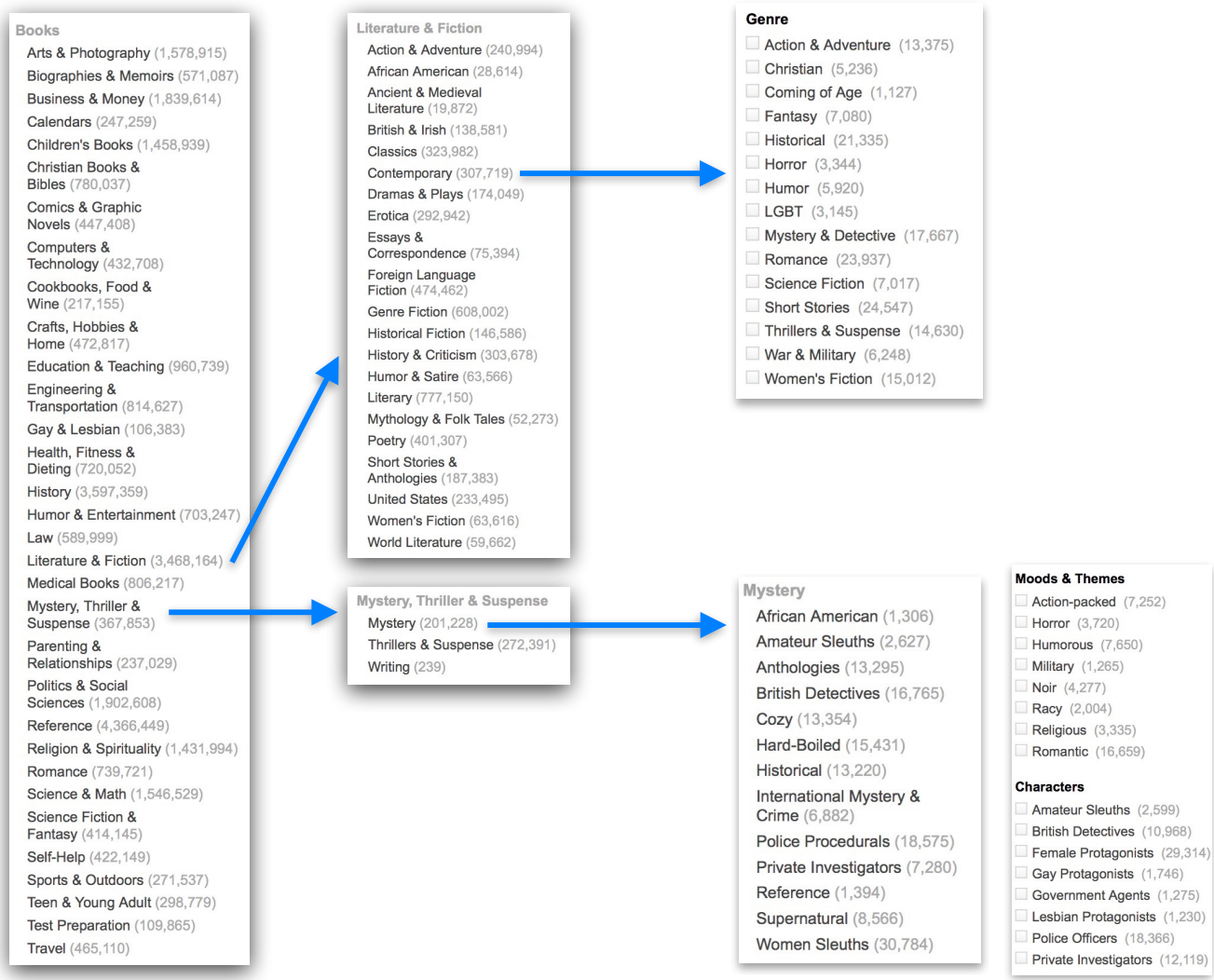


Amazon : Organization of Items (Nodes & Items)

Product Advertising API https://aws.amazon.com/



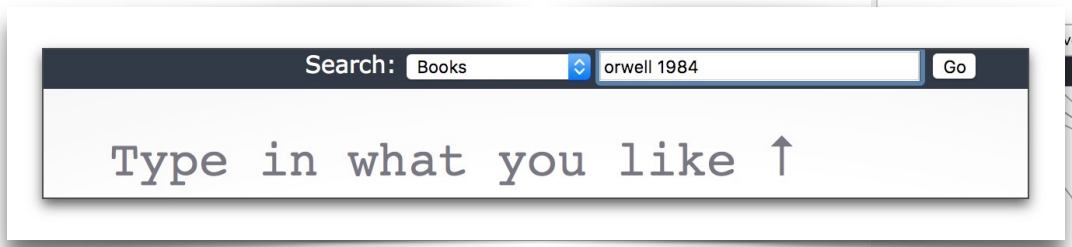

```
<BrowseNodeId>11232</BrowseNodeId>
<Name> Social Sciences</Name>
<Ancestors>
  <BrowseNode>
    <BrowseNodeId>53</BrowseNodeId>
    <Name>Nonfiction</Name>
    <Ancestors>
      <BrowseNode>
        <BrowseNodeId>1000</BrowseNodeId>
        <Name>Subjects</Name>
        <Ancestors>
          <BrowseNode>
            <BrowseNodeId>283155</BrowseNodeId>
            <Name>Books</Name>
          </BrowseNode>
        </Ancestors>
      </BrowseNode>
    </Ancestors>
  </BrowseNode>
</Ancestors>
<Children>
  <BrowseNode>
    <BrowseNodeId>11233</BrowseNodeId>
    <Name>Anthropology</Name>
  </BrowseNode>
  <BrowseNode>
    <BrowseNodeId>11242</BrowseNodeId>
    <Name>Archaeology</Name>
  </BrowseNode>
  <BrowseNode>
    <BrowseNodeId>3048861</BrowseNodeId>
    <Name>Children's Studies</Name>
  </BrowseNode>
</Children>
```



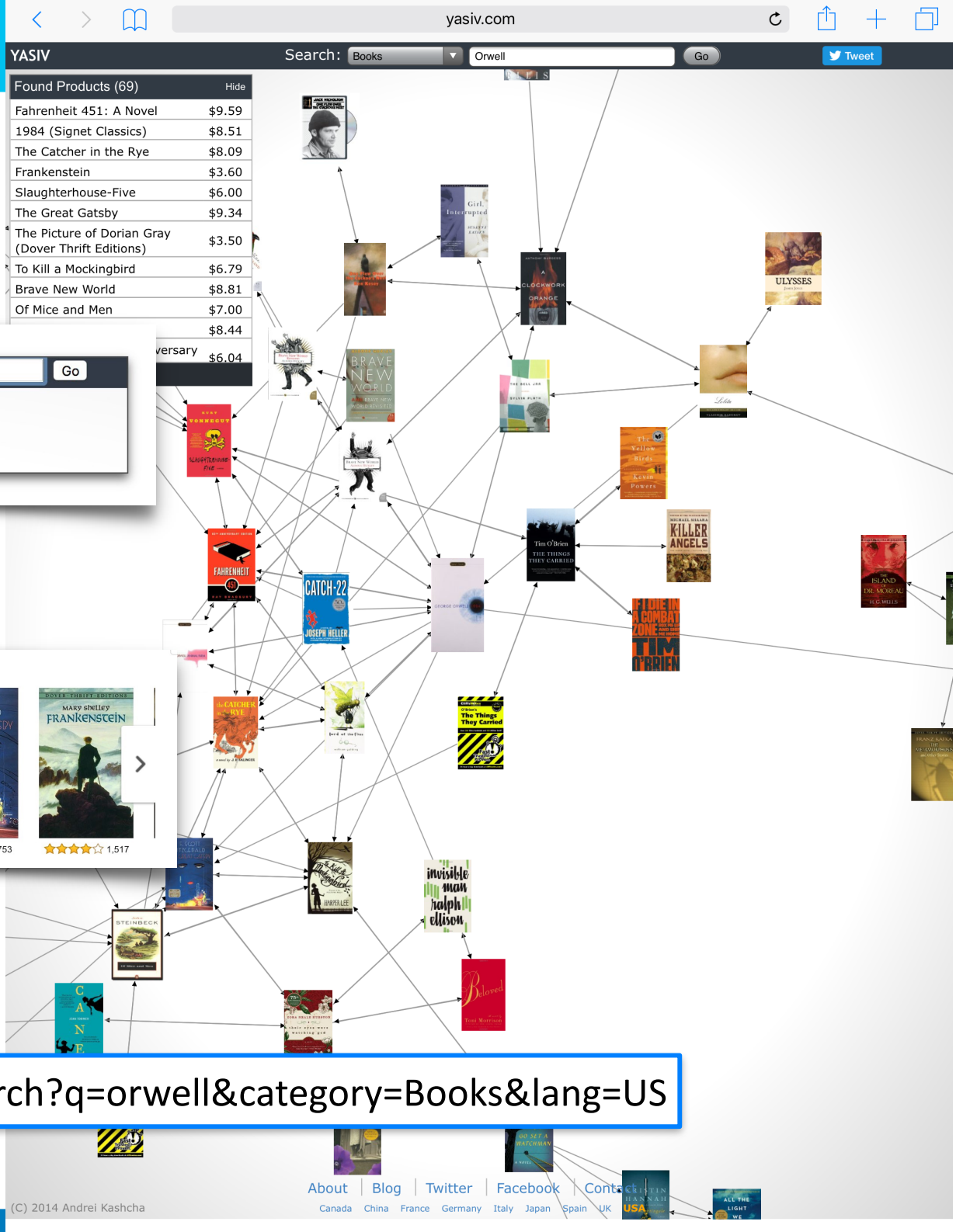
cf. http://www.codediesel.com/libraries/amazon-advertising-api-browsenodes/

Aix*Marseille université

Amazon Navigation Graph : YASIV







http://www.yasiv.com/#/Search?q=orwell&category=Books&lang=US

SBS Collection : Meta-data and Social Data

2.8 Million « Documents » (book descriptions)

Descriptive Information

Social Information

```
−<book>
  <isbn>0001714015</isbn>
  <title>My Book About Me (Beginner Books)</title>
  <ean>9780001714014</ean>
  <binding>Paperback</binding>
  <label>Picture Lions</label>
  <listprice>$10.35</listprice>
  <manufacturer>Picture Lions</manufacturer>
  <publisher>Picture Lions</publisher>
  <readinglevel>Ages 9-12</readinglevel>
  <releasedate/>
  <publicationdate>1983-03-24</publicationdate>
  <studio>Picture Lions</studio>
  <edition/>
  <dewey/>
  <numberofpages>64</numberofpages>
 −<dimensions>
    <height>39</height>
    <width>638</width>
    <length>866</length>
    <weight>35</weight>
  </dimensions>
 −<reviews>
  −<review>
    <date>1996-09-17</date>
    <summary>A FOREVER CHERISHED TREASURE</summary>
   −<content>
      My daughter received this book as a gift for her 5th Birthday. One year later it is
      chocked full of information about her. Each page requires the child to write or draw
      things about their house, school, friends, family, clothes, food, toys, etc. I am surprised
      by the LOW cost of this treasure. It is a great activity book for any age child and once it's
      completed, it provides on-going entertainment. I wish I had this book when I was a child!
      It makes a great gift for any occasion!
    </content>
    <rating>5</rating>
    <totalvotes>4</totalvotes>
    <helpfulvotes>4</helpfulvotes>
  </review>
```

http://social-book-search.humanities.uva.nl

SBS Collection : Real Topics from Library Thing

```
<topic id="812">
  <title>Schumann Biography</title>
  <group>Music History, Performance, and Theory</group>
  <user>PhoenixTerran</user>
  <narrative>
    Does anyone have a recommendation for a biography of Robert Schumann?Thanks!
  </narrative>
  <types>
    <type>subject</type>
  </types>
  <genres>
    <genre>non-fiction</genre>
  </genres>
</topic>
```

```
<topic id="1116">
    <title>Which LISP?</title>
    <mediated_query>introduction book to Lisp</mediated_query>
    <group>Purely Programmers</group>
    <narrative> It'll be time for me to shake things up and learn a new language soon. I had started on Erlang a while back and
           getting back to it might be fun. But I'm starting to lean toward Lisp--probably Common Lisp rather than Scheme. Anyone
           care to recommend a good first Lisp book? Would I be crazy to hope that there's one out there with an emphasis on using
            Lisp in a web development and/or system administration context? Not that I'm unhappy with PHP and Perl, but the best
           way for me to find the time to learn a new language is to use it for my work...
    </narrative>
    <feedback_tags>['artificial intelligence', 'Computing', 'Computers', 'non-fiction', 'ai', 'Reference', 'computer science', '
           programming', 'programming languages', 'Computer programming', 'lisp', 'artificial intelligence', 'ai', 'Reference', '
           computer science', 'Computing', 'own', 'wishlist', 'cs', 'commonlisp', 'Emacs', 'Emacs']
    </feedback_tags>
</topic>
```

http://social-book-search.humanities.uva.nl

— ***Using NLP : to what extent ?*** *(Named Entity Recognition, Syntactic Analysis…)*

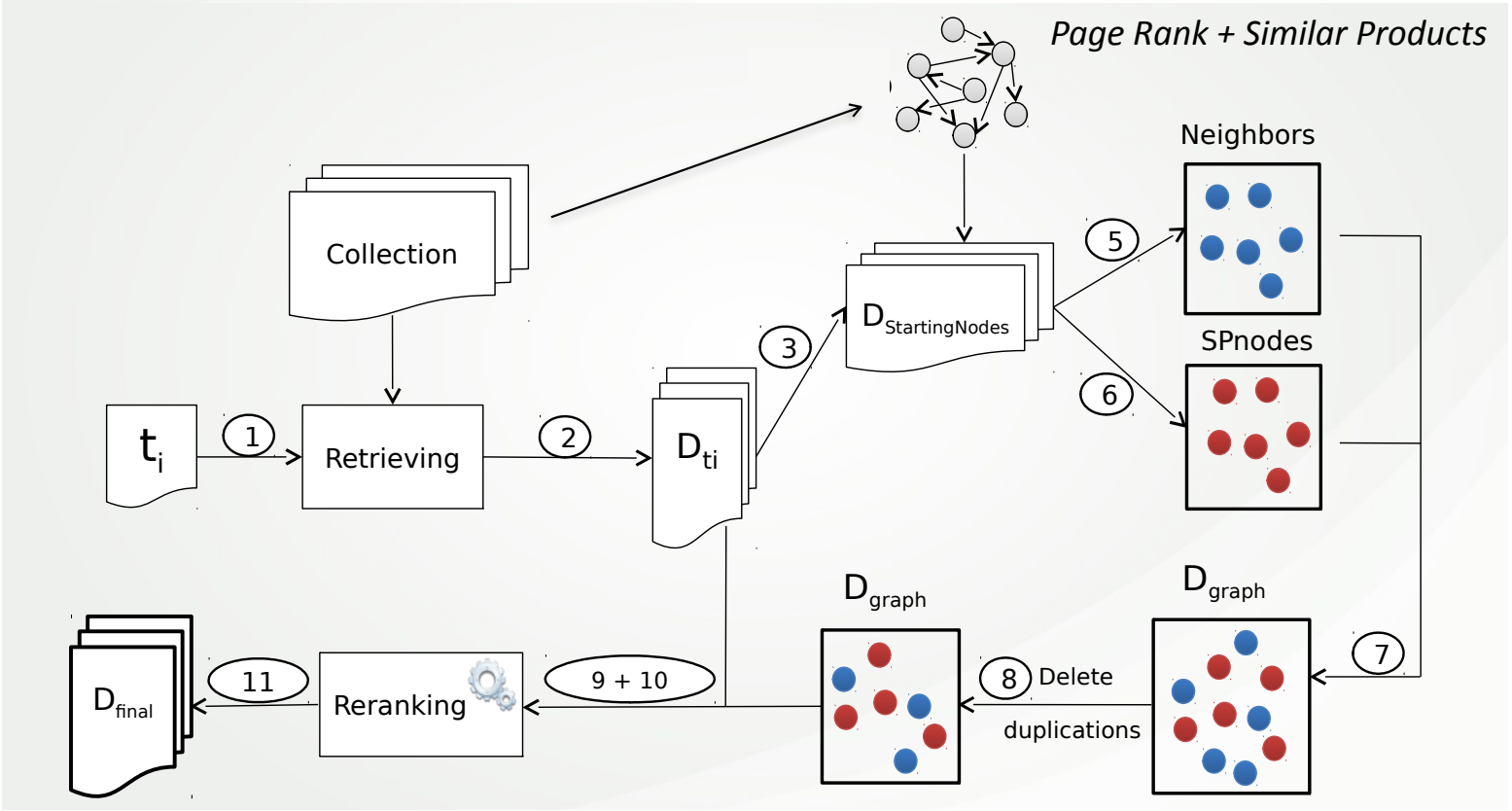
Our proposal for CLEF Social Book Search (Amazon)

IR : Sequential Dependance Model (SDM) - Markov Random Field (Metzler & Croft, 2004) and/or Divergence From Randomness (InL2) model + Query Expansion with Dependance Analysis

Ratings : The more a book has reviews and the more it has good ratings, the more relevant it is.

Graph : Expanding the retrieved books with Similar Books then Reranking with PageRank

$$Likeliness(D) = \log(\#reviews(D)) \times \frac{\sum_{r \in R_D} r}{\#reviews(D)}$$



- Best results in 2011
(Judgements obtained by crowdsourcing)
(IR and ratings)
P@10 ≈ 0.58

- Good results in 2014
(IR, ratings, expansion)
P@10 ≈ 0.23 ; MAP ≈ 0.44

- in 2015 : rank 25/47
(IR + graph
but graph improved IR)
P@10 ≈ 0.2
(best 0.39, included
the price of books)

Our proposal for Scholarly Book Search (OpenEdition)

Langouet, G., (1986), « Innovations pédagogiques et technologies éducatives », Revue française de pédagogie, n° 76, pp. 25-30.

18 Voir Permanent Mandates Commission, *Minutes of the Fifteenth Session* (Geneva: League of Nations, 1929), pp. 100-1. Pour plus de détails, voir Paul Ghali, *Les nationalités détachées de l'Empire ottoman à la suite de la guerre* (Paris: Les Éditions Domat-Montchrestien, 1934), pp. 221-6.

ils ont déjà édité trois recueils et auxquelles ils ont consacré de nombreux travaux critiques. Leur nouvel ouvrage, intitulé *Le Roman véritable. Stratégies préfacielles au XVIIIe siècle* et rédigé à six mains par Jan Herman, Mladen Kozul et Nathalie Kremer – chaque auteur se chargeant de certains chapitres au sein

↓

BILBO →

↓

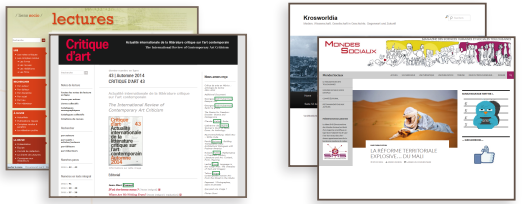
```
<bibl><author><surname>Langouet</surname>,
<forename>G.</forename>,</author> (<date>1986</date>),
<title level="a">« Innovations pédagogiques et
technologies éducatives »</title>, <title
level="j">Revue française de pédagogie</title>,
<abbr>n°</abbr> <biblScope
type="issue">76</biblScope>, <abbr>pp.</abbr>
<biblScope type=page>25-30</biblScope>. <idno
type="DOI">DOI : 10.3406/rfp.1986.1499</idno></bibl>
```

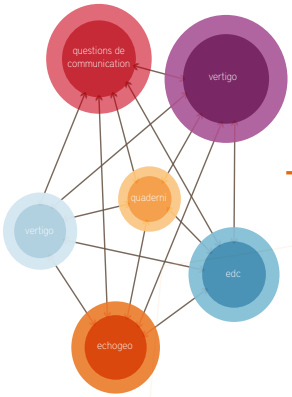
↓

Langouet, G., (1986), « Innovations pédagogiques et technologies éducatives », Revue française de pédagogie, n° 76, pp. 25-30.
DOI : 10.3406/rfp.1986.1499

crossref

OpenEdition

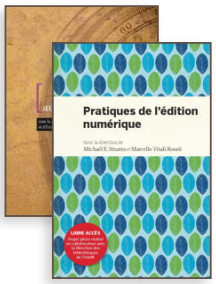




→ RECOMMENDATION

FINDING REVIEWS
LINKING TO BOOKS

GRAPH



RATINGS
POLARITY



INFORMATION EXTRACTION
REFERENCE ANALYSIS

SENTIMENT ANALYSIS

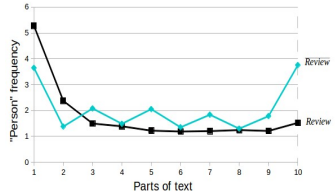
SVM - Z-score - CRF

BOOK RECOMMENDATION

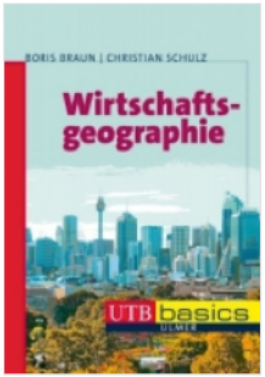
Graph scoring

Searching for Book Reviews in Scientific Journals

- Supervised approaches for filtering = a new kind of genre classification
- Feature selection : Threshold (Z-score) + random forest
- Features : unigrams (words), localisation of named entities / dates

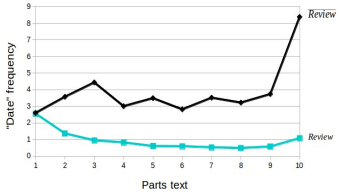


”Person” named entity distribution

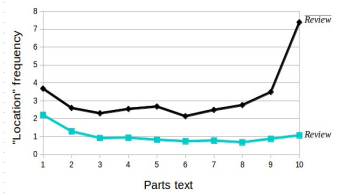


Full size image
Credits : © UTB GmbH

To come straight to the point, the new textbook *Wirtschaftsgeographie* [Economic Geography] by Boris Braun and Christian Schulz is a highly recommendable read. In pursuit of its goal to outline *"a very dynamic and multifarious sub-domain of geography"* [author's translation] (p. 249), this book is mainly directed at current and prospective students of the field of economic geography (p. 6). It keeps both these promises in quite an appealing manner. Alongside an insightful introduction to the myriad of theoretical structures applied in economic geography, its strength lies mainly in its engaging and stylistically coherent review as well as in the systematic cross-linkage of major topics. Valuable didactic tricks – such as info boxes giving a quick overview at the beginning of each chapter, additional text boxes that provide examples to underline the theoretical structures discussed, highlighted keywords in the body of the text, a rich stock of illustrations and charts, as well as review questions and recommendations for further reading – make this book a helpful companion for students. Topics are successfully cross-linked through clearly arranged and colour coded references of lexical precision as well as by repeatedly addressing, relating, and embedding major theoretical approaches from a variety of perception angles. The book embraces the exacting standards

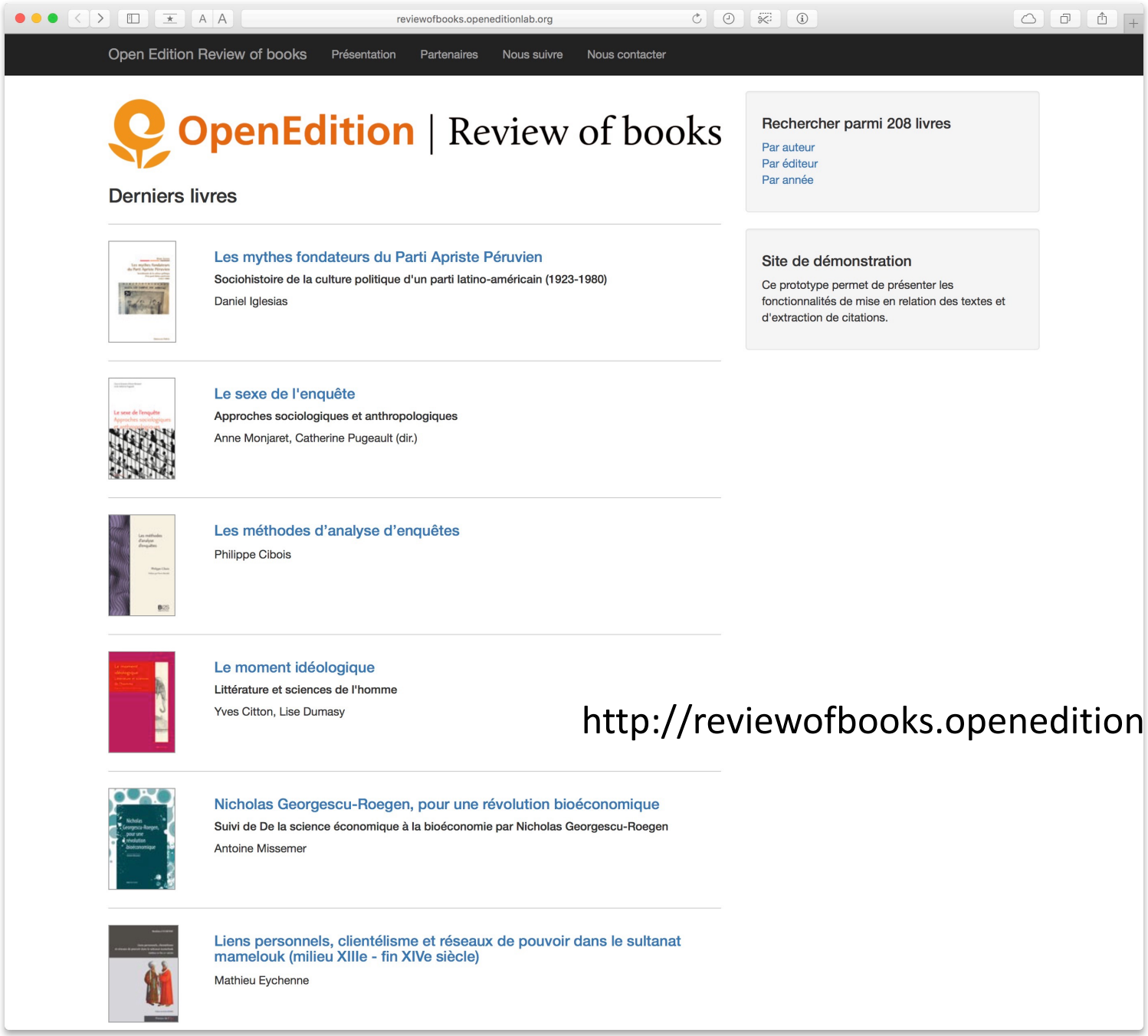


”Date” named entity distribution



”Location” named entity distribution

#	Model	$Review$ R	$Review$ P	$Review$ F-M	$\overline{Review}$ R	$\overline{Review}$ P	$\overline{Review}$ F-M
1	NB	65.5%	81.5%	72.6%	81.6%	65.7%	72.8%
	SVM (Linear)	**99.6%**	**98.3%**	**98.9%**	97.9%	99.5%	98.7%
	SVM (RBF)	89.8%	97.2%	93.4%	96.8%	88.5%	92.5%
	* $C = 5.0$						
	* $\gamma = 0.00185$						
2	NB	**90.6%**	64.2%	75.1%	37.4%	76.3%	50.2%
	SVM (Linear)	87.2%	81.3%	84.2%	75.3%	82.7%	78.8%
	SVM (RBF)	87.2%	**86.5%**	**86.8%**	83.1%	84.0%	83.6%
	* $C = 32.0$						
	* $\gamma = 0.00781$						
3	NB	80.0%	68.4%	73.7%	54.2%	68.7%	60.6%
	SVM (Linear)	77.0%	**81.9%**	79.4%	78.9%	73.5%	76.1%
	SVM (RBF)	**81.2%**	48.6%	**79.9%**	72.6%	75.8%	74.1%
	* $C = 8.0$						
	* $\gamma = 0.03125$						



http://reviewofbooks.openeditionlab.org

Région Provence-Alpes-Côte d'Azur

Linking Contents by Analyzing the References

In Books : no common stylesheet (or a lot of stylesheets poorly respected…)

Our proposal :

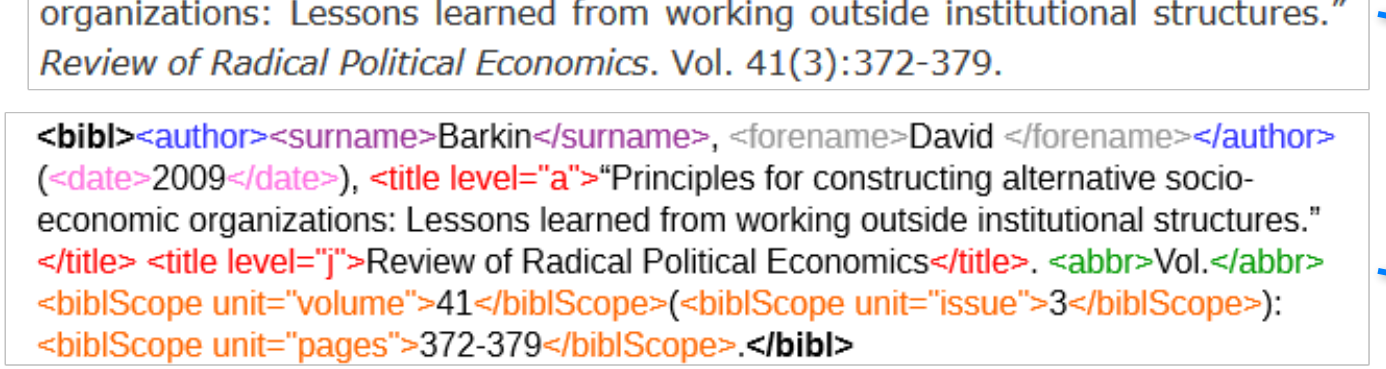
1) Searching for references in the document / footnotes (Support Vector Machines)

2) Annotating the references (Conditional Random Fields)

BILBO : Our (open-source) software for Reference Analysis

Test : http://bilbo.openeditionlab.org
Sources : http://github.com/OpenEdition/bilbo

Barkin, David (2009), "Principles for constructing alternative socio-economic organizations: Lessons learned from working outside institutional structures." *Review of Radical Political Economics*. Vol. 41(3):372-379.



Annotation

Barkin, David (2009), "Principles for constructing alternative socio-economic organizations: Lessons learned from working outside institutional structures." *Review of Radical Political Economics*. Vol. 41(3):372-379.
DOI : 10.1177/0486613409336352

DOI search (Crossref)

OpenEdition Journals : more than 1.5 million references analyzed

Google Digital Humanities Research Awards (2012)



Test : http://bilbo.openeditionlab.org

Sources : http://github.com/OpenEdition/bilbo

Sentiment Analysis on Reviews

- Statistical Metrics (PMI, Z-score, odd ratio…)
- Combined with Linguistic Ressources

Z_score for each term t_i in a class Cj (t_{ij}) $Z_{score(t_{ij})} = \frac{tfr_{ij} - mean_i}{sdi} = \frac{tfr_{ij} - n_j * p(ti)}{\sqrt{nj * p(ti) * (1 - p(ti))}}$

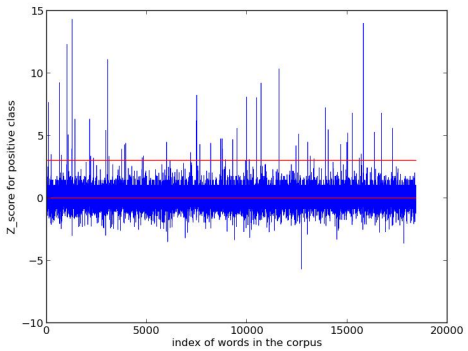


Figure 1 Z_score distribution in positive class

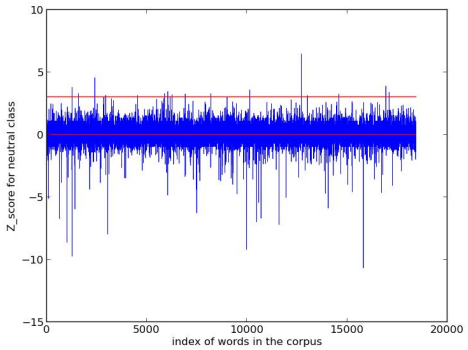
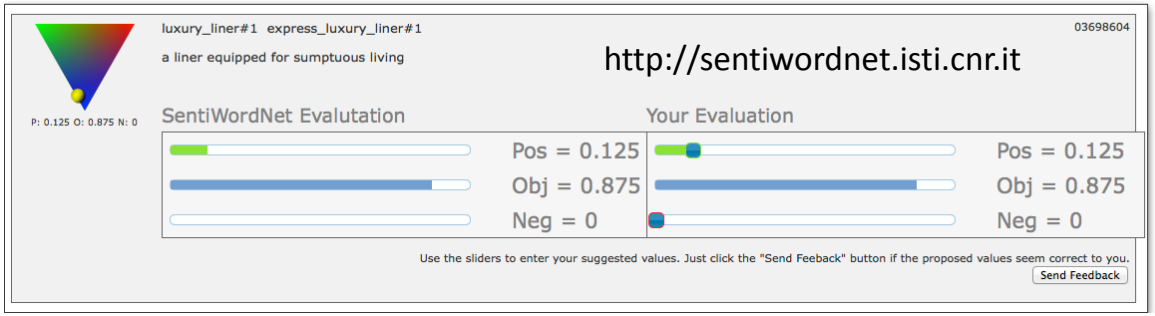


Figure 2 Z_score distribution in neutral class



SentiWordNet

positive	Z_score	negative	Z_score	Neutral	Z_score
Love	14.31	Not	13.99	Httpbit	6.44
Good	14.01	Fuck	12.97	Httpfb	4.56
Happy	12.30	Don't	10.97	Httpbnd	3.78
Great	11.10	Shit	8.99	Intern	3.58
Excite	10.35	Bad	8.40	Nov	3.45
Best	9.24	Hate	8.29	Httpdlvr	3.40
Thank	9.21	Sad	8.28	Open	3.30
Hope	8.24	Sorry	8.11	Live	3.28
Cant	8.10	Cancel	7.53	Cloud	3.28
Wait	8.05	stupid	6.83	begin	3.17

Table1. The first ten terms having the highest Z_score in each class

Features	F-measure	
	2013	2014
Terms	50.15	48.56
Terms+Z	57.17	58.37
Terms+POS	44.07	42.64
Terms+POL	54.72	54.53
Terms+Z+POS	53.20	56.47
Terms+Z+POL	**59.66**	**61.07**
Terms+POS+POL	48.97	51.90
Terms+Z+POS+POL	55.83	60.22

Table 3. Average f-measures for positive and negative classes of SemEval2013 and 2014 test sets after using a twitter dictionary.

[Hamdan, Béchet & Bellot, SemEval 2014]

Aix*Marseille université



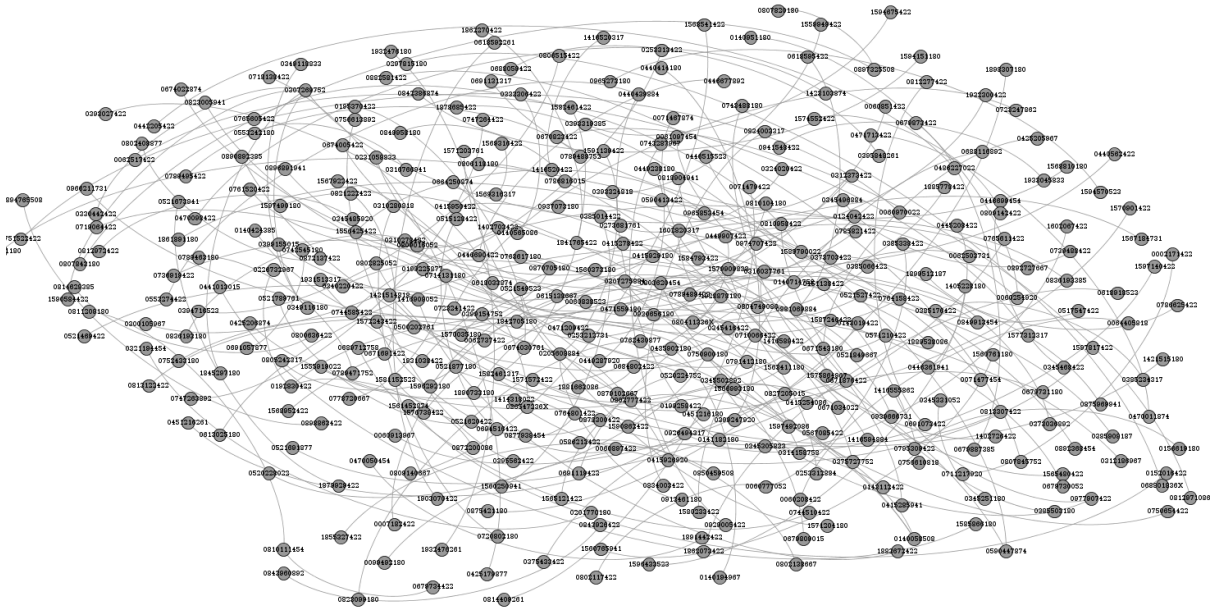
Building a Graph of Books

— Nodes = books + properties (metadata, #reviews and ranking, page ranks, ratings…)

— Edges = links between books

  — Book *A refers to Book B* according to:
    — Bibliographic references and citations (in the book / in the reviews)
    — Amazon recommendation (*People who bought A bought B, People who liked A liked B…*)

  — *A is similar to B*
    — They share bibliographic references
    — Full-text similarity + similarity between the metadata



The graph allows to estimate
— « Book Ranks »
(cf. the Google's Page Rank)
— Neighborhood
— Shortest paths

…Evaluation in progress…

Conclusion : (Still) (Half)Open Questions

— How to combine Information Retrieval and Recommendation ?
  - Learning to Rank : the best way now
  but… what is a good recommendation is not clear (crowdsourcing ?)
  - Using Reader Profiles (cf. CLEF Social Book Search track since 2015)

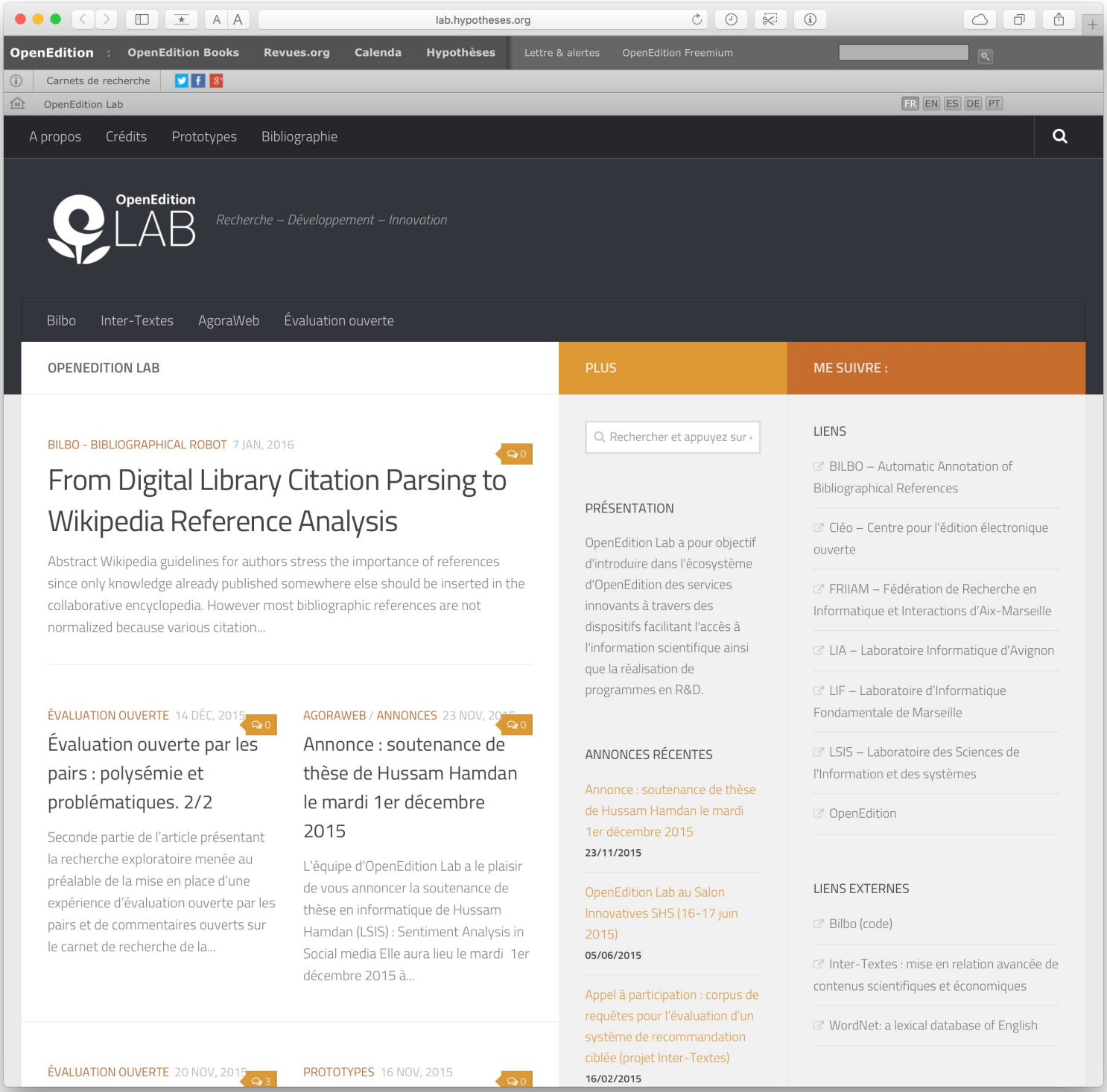
— Is combining « official » meta-data and Reader-generated content (reviews) useful ?
  YES

— How to go beyond topical relevance ? (quality, authority, genre, expertise…)
  Combining Graph analysis / Sentiment analysis
  Natural Language Processing

— How to deal with complex queries ?
  NLP (dependencies, sub-queries, aspects…) — cf. *Question Answering*
  Translating Natural Language Query into Structured Queries

— How to identify and exploit references (cross-linking) ?
  Supervised probabilistic approaches (robust, effective and efficient)

— Is the content of the books useful ? (vs. reviews / abstracts only) ?
  Not sure for retrieval…
  For linking : YES (at least when the book is not a best-seller - not many reviews)



http://lab.hypotheses.org