

A Question Answering service for information retrieval in Cooper

Citation for published version (APA):

Giesbers, B., Van der Vegt, W., Van Bruggen, J., Koper, R., & Taddeo, A. (2008). A Question Answering service for information retrieval in Cooper. In S. Ceri, W. Nejdl, J. van Bruggen, & F. Van Assche (Eds.), *Proceedings of the 1st International Workshop on Collaborative Open Environments for Project-Centered Learning* (Vol. 309, pp. 76-85). CEUR-WS.org. <http://ceur-ws.org/Vol-309/>

Document status and date:

Published: 07/01/2008

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 10 Oct. 2024

Open Universiteit
www.ou.nl



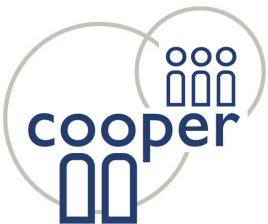


Collaborative Open Environment for Project Centered Learning

A Question Answering service for information retrieval in Cooper

Bas Giesbers (OUNL), Antonio Taddeo (ALaRI), Wim van der Vegt (OUNL), Philip Liesefeld (CoWare)
Jan van Bruggen (OUNL), Rob Koper (OUNL)

- What do we want with QA in Cooper?
- Latent Semantic Analysis
 - Technique
 - Demonstration
 - Assumptions/requirements
- Toolbox
 - Architecture
- Future assessment in Cooper
 - Implementation at ALaRI
 - Implementation at CoWare
- Conclusion, Discussion and questions



What do we want with QA in Cooper?

- Support of (standard) activities in Cooper.
- Relate documents to people, places and things.
- Solution: a search engine.
- But, lexical methods won't suffice:
 - Changing information needs
 - Context dependency
 - Small, specific domains
- Now what?



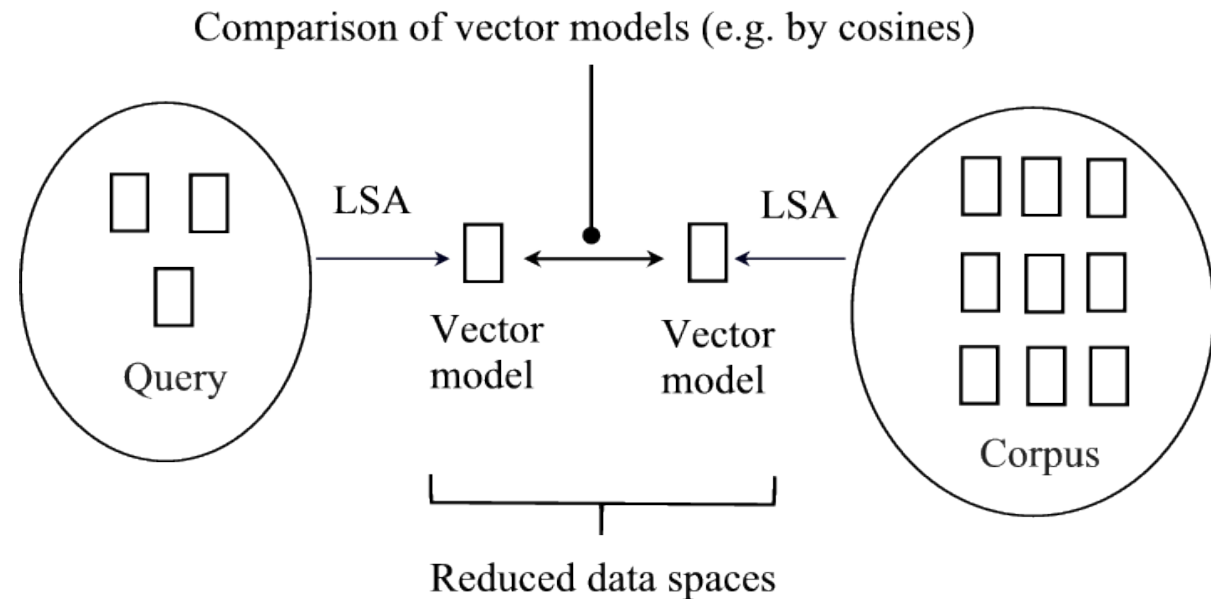
Latent Semantic Analysis

- ‘...a computational method by which a major component of language learning and use can be achieved.’ (Landauer, 2007, p.10).
- Used in:
 - Cognitive science (memory research)
 - Education (assessment, automatic tutoring)
 - Information retrieval
- An analysis
 - Turn collection of documents into Term x Document matrix
 - Perform SVD on the matrix
 - Reduce the material , i.e. remove smallest singular values
 - Reproduce original
 - Represent document by vector
 - Calculate cosines between vectors



Latent Semantic Analysis (2)

- LSA explains the content of a text as the weighted sum of underlying constructs.
 - Similar to Factor Analysis and PCA
- Query:



Demonstration

- Domain specific corpus: combination of CoWare, OUNL and L3S documents (300 total).
- For good mathematical representation, corpus should be large enough to let the machine ‘learn the language’.
 - TASA as additional material
- Pre-processing
 - Convert to ASCII txt
 - If first line empty, remove
 - Remove certain diacritical marks
- Software
 - **GTP** (<http://www.cs.utk.edu/~lsi/soft.html>)
 - **R** (<http://www.r-project.org/>)
 - **Infomap** (<http://infomap-nlp.sourceforge.net/>)



Demonstration (2)

- Query: Three documents from OUNL repository:
 - IMS LD introduction
 - IMS LD pedagogy
 - Building Blocks for a Smart Space for Learning
- Results:
 - Documents with same terms are related (OUNL - OUNL)
 - Term doesn't have to occur to find relation (OUNL – L3S)
 - Term can occur but no relation is found (OUNL - CoWare)
 - Threshold: which docs to report and which not?



Demonstration (3)

Microsoft Excel - EcTel_highlights

Bestand Bewerken Beeld Invoegen Opmaak Extra Data Venster Help Adobe PDF

Typ een vraag voor hulp

Files in Corpus

	A	B	C	D	E
1	Files in Corpus	BuildingBlocks	IMSLDped	IntroIMSLD	
2	M:\Coware\corpus\103061383336021013052802247904608673059.txt 0	0,50	0,81	0,93	
3	M:\Coware\corpus\107926078625827310826448553703965214491.txt 0	0,59	0,80	0,94	
4	M:\Coware\corpus\108458902078808154216869291613774511857.txt 0	0,06	0,07	0,07	
5	M:\Coware\corpus\108511470198353383228937243108063448692.txt 0	0,64	0,71	0,89	
6	M:\Coware\corpus\109566258988385654540660081919832501472.txt 0	0,69	0,78	0,87	
7	M:\Coware\corpus\109590769831402879110745721832672523348.txt 0	0,01	0,03	0,02	
8	M:\Coware\corpus\109779521259383397233732951204798029149.txt 0	0,57	0,67	0,91	
9	M:\Coware\corpus\11045793391740141687391585879149161363.txt 0	0,64	0,75	0,95	
10	M:\Coware\corpus\111919110645380868090129685634032098155.txt 0	0,51	0,85	0,82	
11	M:\Coware\corpus\112435728632733658105503004606457311091.txt 0	0,66	0,69	0,87	
12	M:\Coware\corpus\114230135429661530193768680013103504081.txt 0	0,44	0,59	0,48	
13	M:\Coware\corpus\11459810015766233438401785180152051198.txt 0	0,64	0,72	0,94	
14	M:\Coware\corpus\114606902619171854911656745545066544104.txt 0	0,44	0,61	0,68	
15	M:\Coware\corpus\114698227628972558543125678449081928775.txt 0	0,58	0,87	0,95	
16	M:\Coware\corpus\Learning%20Objects%20and%20the%20Semantic%20We	0,71	0,70	0,87	
17	M:\Coware\corpus\lecture_a_instruction_set.txt 0	0,08	0,10	0,06	
18	M:\Coware\corpus\lecture_b_pipeline.txt 0	0,08	0,09	0,05	
19	M:\Coware\corpus\lecture_c_hazard.txt 0	0,02	-0,01	0,02	
20	M:\Coware\corpus\lecture_c_hazard2.txt 0	0,02	-0,01	0,02	
21	M:\Coware\corpus\Lessons%20Learned%20in%20Designing%20a%203D%20	0,40	0,33	0,23	
22	M:\Coware\corpus\logical_characterization_henze_nejdl.txt 0	0,37	0,44	0,26	
23	M:\Coware\corpus\mix-vldb2005.subm.txt 0	0,26	0,12	0,06	
24	M:\Coware\corpus\Modeling%20and%20Evaluation%20of%20Scatternets%20	0,38	0,24	0,24	
25	M:\Coware\corpus\navigationengforsemanticweb.txt 0	0,41	0,16	0,19	
26	M:\Coware\corpus\Nejdl_Wolf_et_al-EDUTELLA_Searching_and_Annotating_	0,76	0,21	0,15	
27	M:\Coware\corpus\OAI-P2P%3a%20A%20Peer-to-Peer%20Network%20for%	0,64	0,15	0,11	
28	M:\Coware\corpus\Painter_Brase_IFETJournal.txt 0	0,73	0,36	0,30	
29	M:\Coware\corpus\qu_ijdl.txt 0	0,49	0,41	0,39	
30	M:\Coware\corpus\Qu_ITCC.txt 0	0,33	0,20	0,12	
31	M:\Coware\corpus\Re-Designing%20an%20Educational%20Setting%20-%20	0,70	0,70	0,87	
32	M:\Coware\corpus\Relevance_Haptic_EDMedia2002.txt 0	0,59	0,59	0,66	
33	M:\Coware\corpus\simulation_p2pdb.txt 0	0,36	0,16	0,11	
34	M:\Coware\corpus\SMART_LTB_BG_0001.txt 0	0,16	0,20	0,09	
35	M:\Coware\corpus\SMART_L2C_SL_0001.txt 0	0,15	0,16	0,16	
36	M:\Coware\corpus\Steinke_EDMedia2003.txt 0	0,55	0,57	0,68	
37	M:\Coware\corpus\Steinke_EDMedia2004.txt 0	0,26	0,25	0,28	
38	M:\Coware\corpus\Steinke_FLearn2003.txt 0	0,32	0,31	0,32	

matrix

Gereed

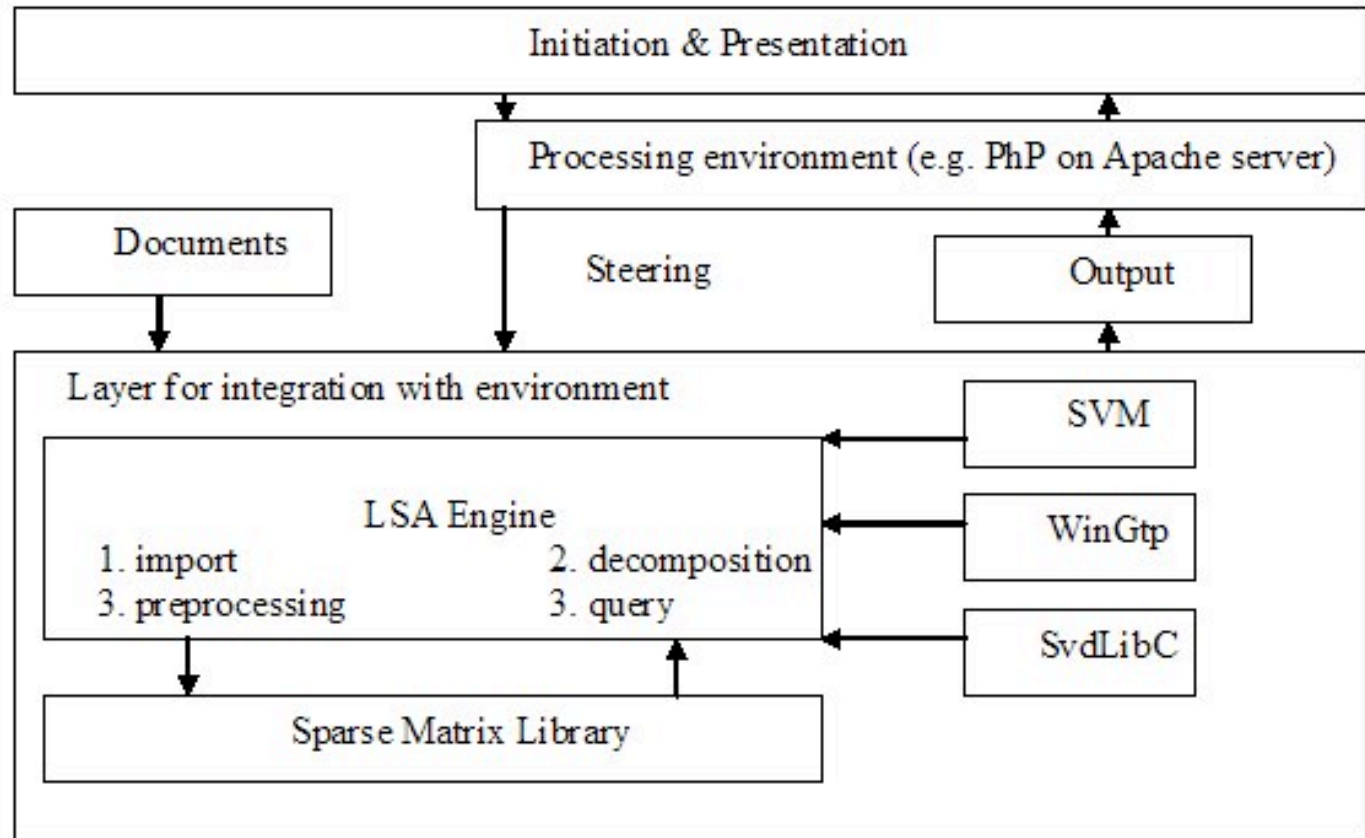
Start Postvak I... coware Google - ... Microsoft ... iTunes Voorbeeld... TextPad -... Adobe Ac... EcTel_hi... NL 11:56

Review

- Relate documents to people, places and things.
- More than lexical methods: Latent *Semantic* Analysis.
- LSA is usable, existing applications may not.
 - Pre-processing and post-processing necessary
 - Requirement to function as a webservice

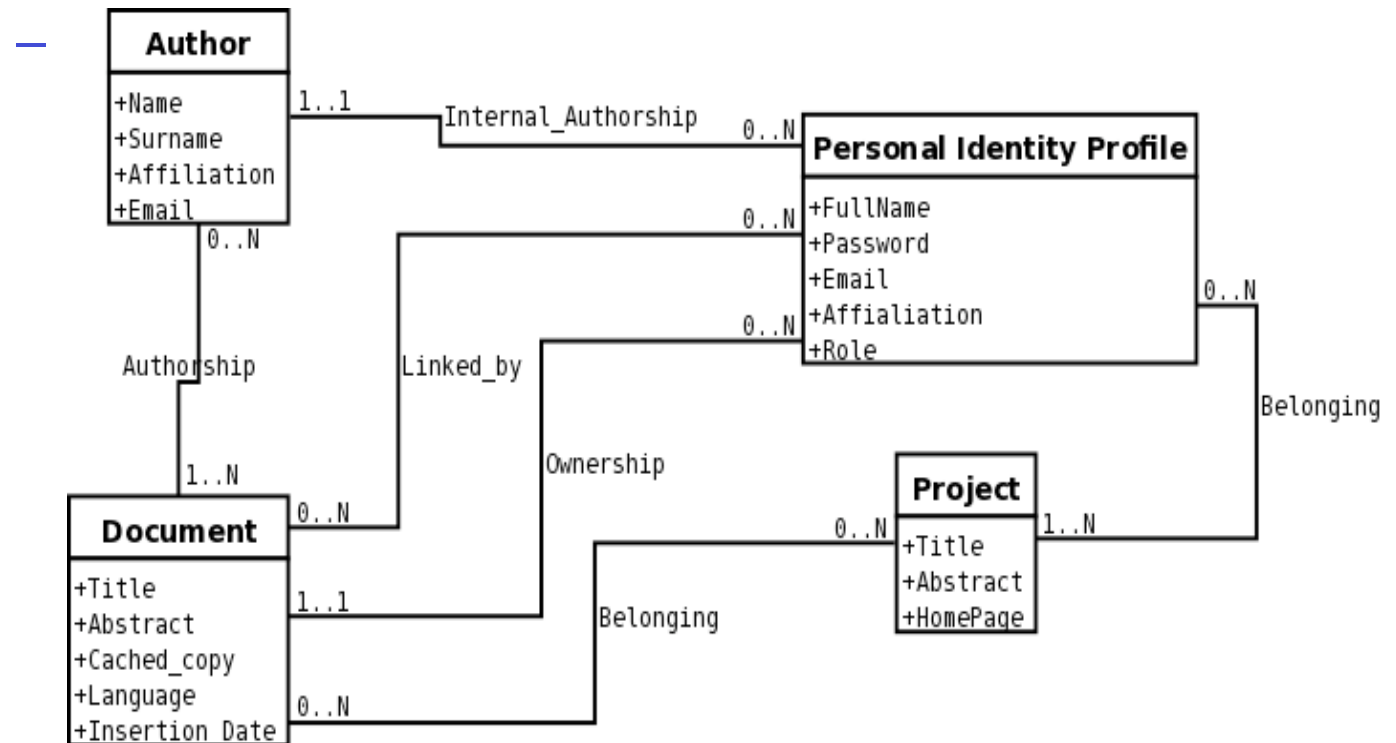


Toolbox and architecture



Future assesement: implementation at ALaRI

- Relate documents to people, places and things.



Future assesement: implementation at ALaRI

– Document Detail page at ALaRI:

The screenshot shows a web browser window displaying the ALaRI Document Detail page. The browser's address bar shows the URL: <http://www.alari.ch/intranet/page1658.do?link=ln7434&sp=page1661&dau2282.oid=270>. The page features a navigation menu on the left with options like 'Intranet Home', 'Change Password', and 'Logout'. The main content area is titled 'Document Detail' and includes the following information:

- Title:** Assessing security in energy-efficient sensor networks
- Abstract:** Abstract In the EYES project (<http://eyes.eu.org>), we are investigating self-organizing, collaborative, energy-efficient sensor networks. This study is devoted to the security aspects of the project. Our contribution is three-fold: firstly, we present a survey, where we discuss the dominant issues of energy-security trade-off in the network protocol and key management design space. From there we set out future research directions for our security framework. Secondly, we propose an assessment framework based on system profile, with which we have managed to carve out manageable design spaces from the seemingly infinite possibilities of ad hoc mobile wireless networks. Finally, we have benchmarked some well-known cryptographic algorithms in search for the best compromise in security and energy-efficiency, on a typical sensor node. Our preliminary investigations also cover an important parameter in the design space: the resource requirements of the symmetric key algorithms RC5 and TEA.
- Comments:** Very good overview about Security-Critical Energy-Efficient Protocols for Sensor Networks
- Type:** Scientific paper
- Location:**
- Year:** 2002
- Date uploaded:** 11/04/2005
- Keywords:**
- Language:** English
- Number of Links:** 3

A 'Download' button is located below the document details. To the right, there are sections for 'Authors' (listing S. Etalle, P. Hartel, and Y. Law), 'Doc Is Bibliography for project' (listing 'System-Level Design of the security concept of a Remote Meter Reader'), 'My Related Virtual Folders' (listing 'Security' and 'Sensor Networks'), and 'Doc Uploaded By...' (listing 'Spanagel Daniel', 'Alumni 04-05').

Below the document details, a red-bordered box highlights the 'Search Results - Whole repository' section, which lists several related documents:

- ▶ A Bluetooth Based Sensor Network for Civil Infrastructure Health Monitoring
- ▶ ASP: An Adaptive Energy-Efficient Polling Algorithm for Bluetooth Piconets
- ▶ Collaborative Robotics with Lego Mindstorms
- ▶ Energy-Efficient Communication Protocol for Wireless Microsensors Networks
- ▶ Middleware to Support Sensor Network Applications
- ▶ Physical Layer Driven Protocol and Algorithm Design for Energy Efficient Wireless Sensor Networks

At the bottom of the page, a 'You can...' section provides links for 'Add to your Project Bibliography', 'Link Doc to Other Projects', and 'Link/UNLink Doc to your Virtual Folders'.

Future assesement: implementation at CoWare

- Find related documents (no people and/or things).
- And: make a distinction between expert and novice documents.
- This is work in progress.



Conclusion and discussion

- QA to support Cooper activities.
- Not lexical, but semantic relations by using LSA.
- Development of toolbox and architecture that can:
 - Pre-process and post-process material
 - Perform complete SVD
 - Can be used as a webservice
- Implementation at ALaRI and CoWare.



Questions

Thank you for your attention!

Any questions?



References

- Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W. (eds.): Handbook of Latent Semantic Analysis (2007). Mahwah, New Jersey: Lawrence Erlbaum Associates
- Macedo, A.A., Pimentel, M.G.C., Cammacho-Guerrero, J.A. (2002). An infrastructure for open latent semantic linking. In: Proc ACM HT'02, pp 107–116
- Iofciu, T., Zhou, X., Giesbers, B., Rusman, E., van Bruggen, J., Ceri, S. (2006). State of the Art Report in Knowledge Sharing, Recommendation and Latent Semantic Analysis.
- Giles, J. T., Wo, L., & Berry, M. W. (2001). GTP (General Text Parser) Software for Text Mining. In Statistical Data Mining and Knowledge Discovery (chap. 27). CRC Press. Retrieved from <http://www.cs.utk.edu/~berry/papers02/GTPchap.pdf>

