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)

Information Society Technologies

Overview

- 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).

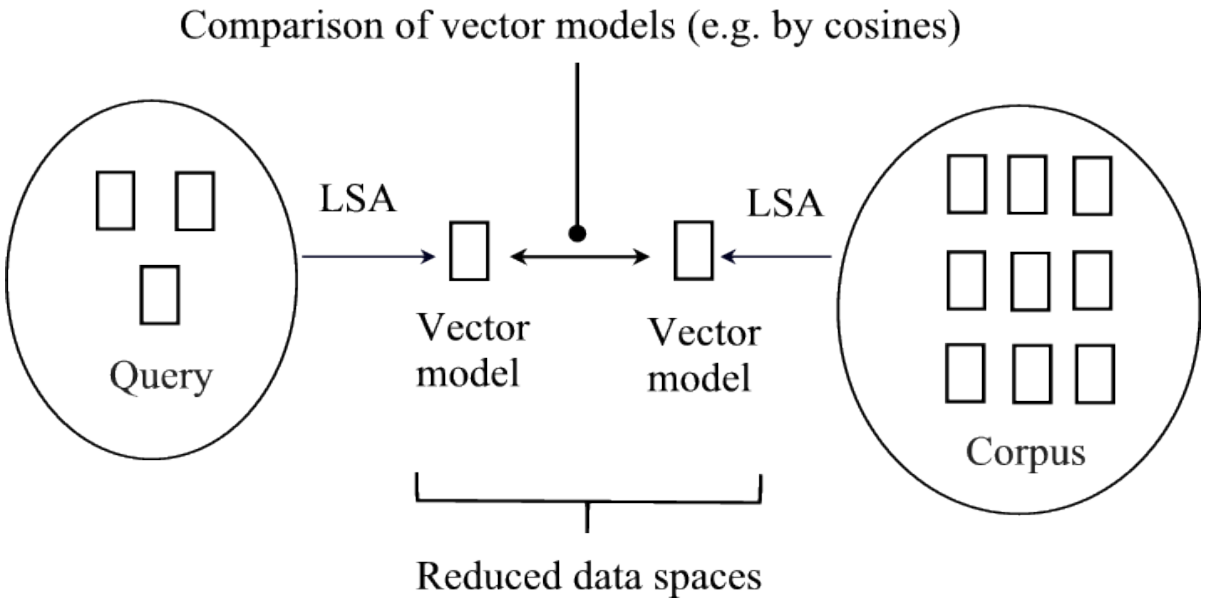
- Used in:
  - Cognitive science (memory research)
  - Education (assessment, automatic tutoring)
  - Information retrieval

- 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:



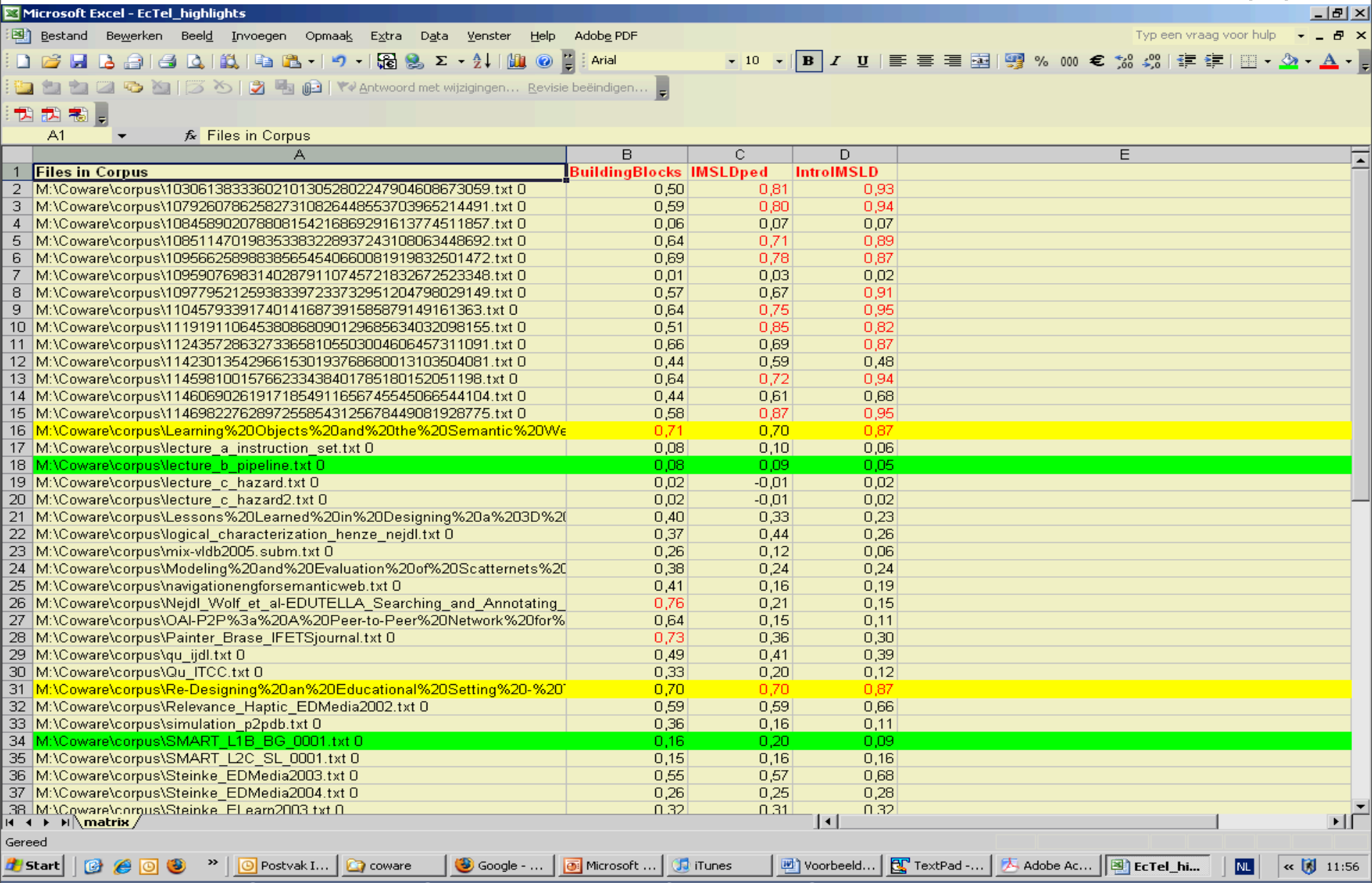
Demonstration

- 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)

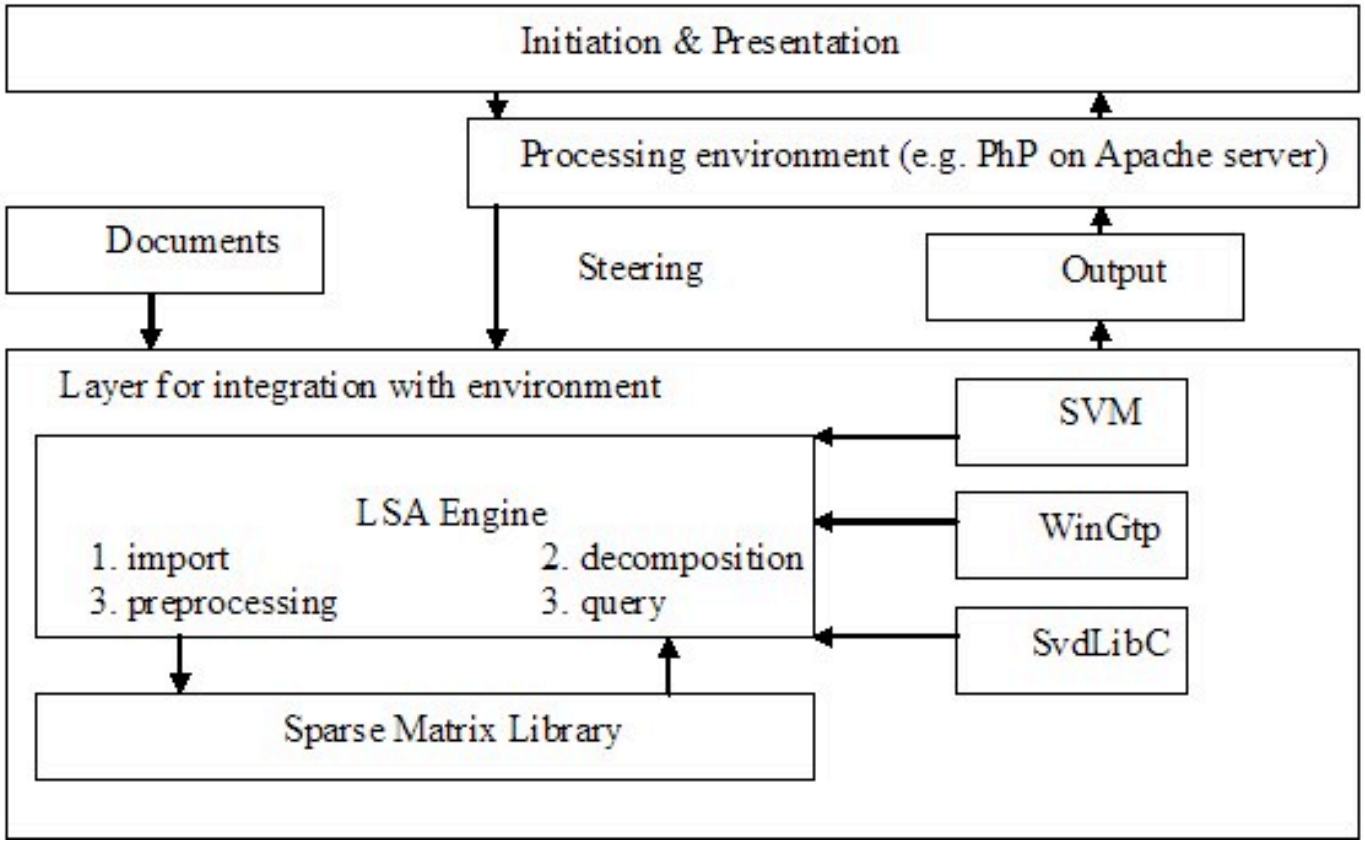


	A	B	C	D	E
1	Files in Corpus	BuildingBlocks	IMSLDped	IntroIMSLD	
2	M:\Coware\corpus\10306138333602101305280224790460867305 9.txt 0	0,50	0,81	0,93	
3	M:\Coware\corpus\10792607862582731082644855370396521 4491.txt 0	0,59	0,80	0,94	
4	M:\Coware\corpus\10845890207880815421686929161377451 1857.txt 0	0,06	0,07	0,07	
5	M:\Coware\corpus\10851147019835338322893724310806344 8692.txt 0	0,64	0,71	0,89	
6	M:\Coware\corpus\10956625898838565454066008191983250 1472.txt 0	0,69	0,78	0,87	
7	M:\Coware\corpus\10959076983140287911074572183267252 3348.txt 0	0,01	0,03	0,02	
8	M:\Coware\corpus\10977952125938339723373295120479802 9149.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\11191911064538086809012968563403209 8155.txt 0	0,51	0,85	0,82	
11	M:\Coware\corpus\11243572863273365810550300460645731 1091.txt 0	0,66	0,69	0,87	
12	M:\Coware\corpus\11423013542966153019376868001310350 4081.txt 0	0,44	0,59	0,48	
13	M:\Coware\corpus\11459810015766233438401785180152051 198.txt 0	0,64	0,72	0,94	
14	M:\Coware\corpus\11460690261917185491165674554506654 4104.txt 0	0,44	0,61	0,68	
15	M:\Coware\corpus\11469822762897255854312567844908192 8775.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%2(	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%2(	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_IFETSjournal.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_L1B_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_ELearn2003.txt 0	0,32	0,31	0,32	

Review

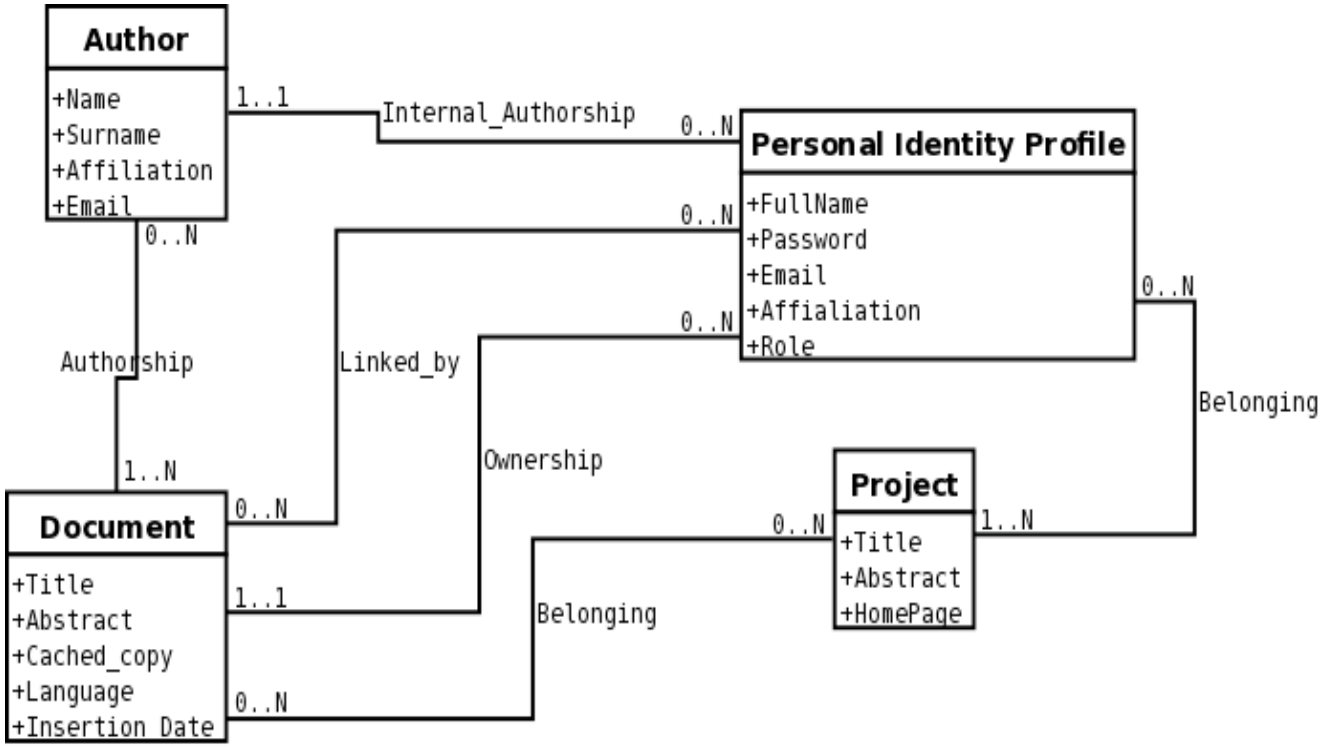
- Relate documents to people, places and things.
- More than lexical methods: Latent *Semantic* Analysis.
- LSA is usable, existing applications may not.
  - Pre-processing and post-processing necessary
  - Requirement to function as a webservice

Toolbox and architecture



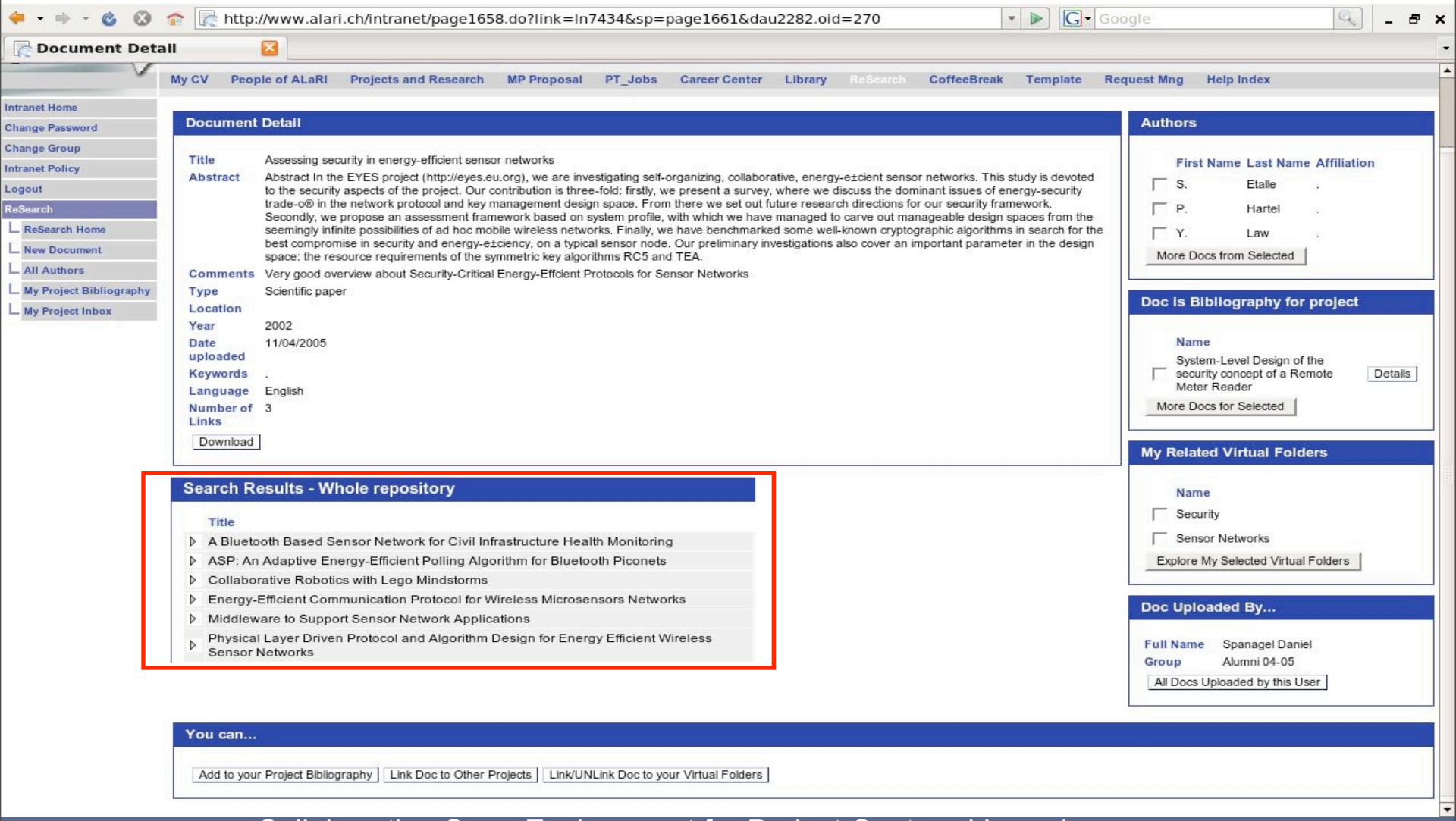
Future assessement: implementation at ALaRI

- Relate documents to people, places and things.
-



Future assessement: implementation at ALaRI

– Document Detail page at ALaRI:



Future assessement: 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