The University Of Sheffield.

Automatic Techniques for Extracting Semantic Data (from text and media)

Professor Fabio Ciravegna
Intelligent Web Technology Lab,
Natural Language Processing Group
Department of Computer Science, University of Sheffield,
Sheffield. United Kingdom
http://www.dcs.shef.ac.uk/~fabio/

Sponsored by
www.3worldt.org

Sponsored by
X-MEDIA
www.x-media-project.org

Knowledge Workers' Challenges

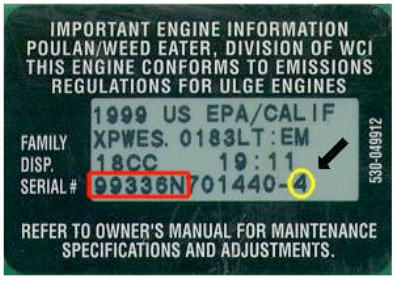
- Gathering knowledge relevant to a task or problem
  - it may be distributed across different storage systems and different media
- Analysing the knowledge they have gathered and making sense of it
- Sharing knowledge with their colleagues
- Keeping track of the process
  - by being aware of what one is doing, what one needs to do next, and what others are doing
- What to search for, what analysis is needed and who to share with
  - depend on the task in hand and the current stage of the process

Uren *et al.* 2007: X-Media Deliverable 4.1: Specification of Knowledge Sharing Systems

An Example of Knowledge Management

jet engines are moving towards complete serialisation
- every piece has a serial number (excepts nuts and bolts)
- the history of each part is recorded
  - e.g. part robbed to engine

© Rolls-Royce plc



99336N = Date Code
- Day of the Year
- Year of Production

4 = Product Type

Jet engine example

- a jet engine can produce ~1Gbyte of vibration data per hour of flight;
  - if irregularities are found, part of the data can be stored
  - reports can be written (event reports)
  - pictures can be taken

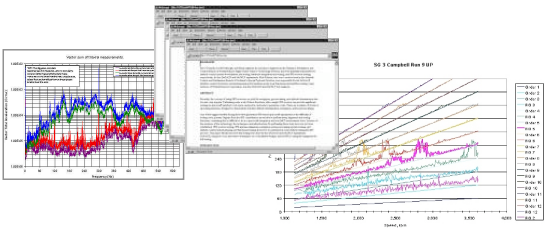
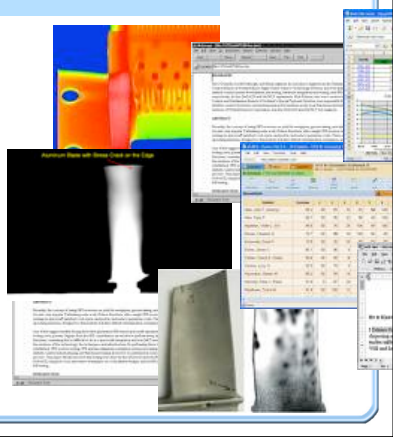


image © Rolls-Royce plc

Jet engine example (3)

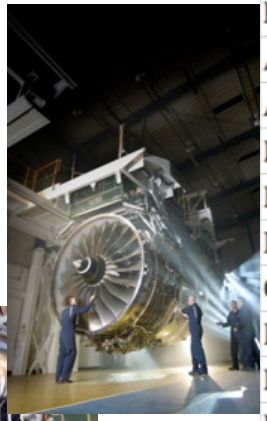
When engine is serviced (e.g. overhaul)

- financial information is produced.
- if problems are found,
  - pictures are taken
  - reports are written
  - engine is tested



Jet engine example (4)

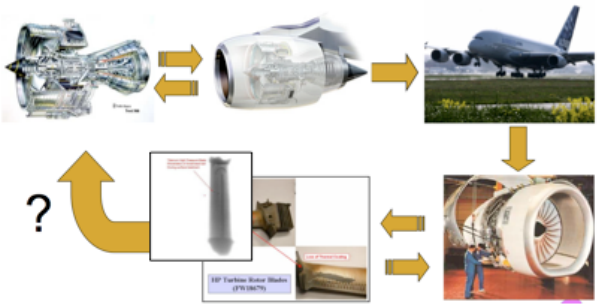
- If issue is recurring (or suspected so)
  - an issue resolution group is established
    - existing evidence is retrieved
    - further evidence is collected
    - a learned lesson is generated
    - same issues is investigated across models

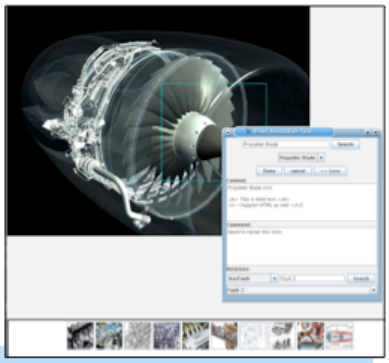


Document Type
AROC proforma
AROC results
Development
EHM data
Emails
ONWING emails
Images
Lab findings
Monitoring Requir
Presentations
Procedures
RCP
Risk Assessment
Solution Reports
Technical Reports
TS&O Reports

Jet engine example (4)

- Lifecycle "folder" will easily sum up to several Terabytes
- Folder will contain highly interrelated information stored in different media





- Goal for Knowledge Management:
- Making information available independently from
  - Data format (structured/unstructured)
  - The archive
- Making it available for automatic processing
- Making it easily accessible and manageable despite its size

The University Of Sheffield.

What do we know and what we do not

- As we know, there are known knowns
  - that are things we know we know.
- We also know there are known unknowns;
  - that is to say we know there are some things we do not know.
- But there are also unknown unknowns
  - the ones we don't know we don't know

Donald Rumsfeld

The University Of Sheffield.

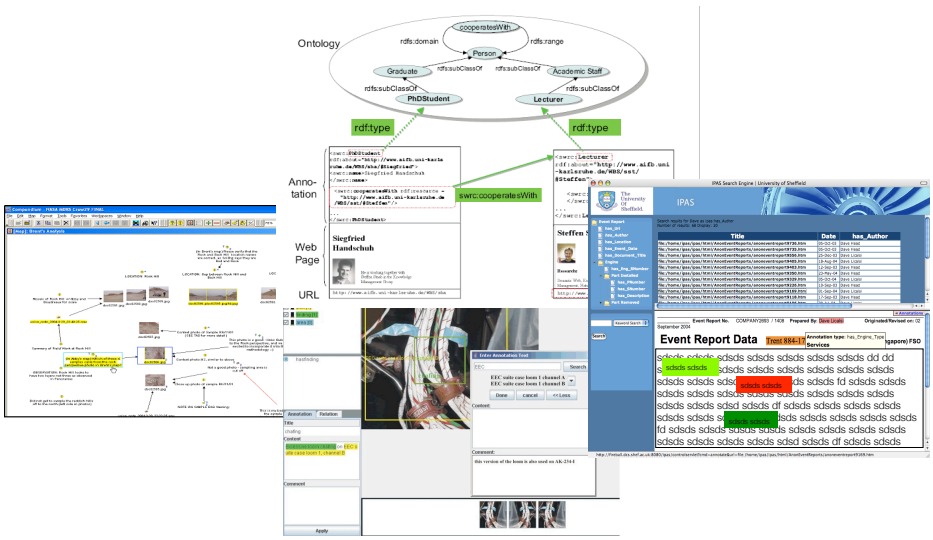
Failing factors: Technical Issues

- Information scattered in multiple repositories
  - No one really knows which information is available and/or where
  - There isn't a single access point to information
  - Even a company-wide keyword searching facility is often inexistent
- 80-85% of a company's knowledge is unstructured
  - i.e. expressed in some forms of natural language or images/videos
- Information overload
  - Growing archives
  - Cost of storing very low
    - Video and 2D/3D image storing a reality

The University Of Sheffield.

Management of What Type of Knowledge?

- Internal Knowledge (often on a *very large* Web Intranet -- millions of pages)
  - Need: capturing and sharing
  - e.g. How to design a product
- Focused external knowledge (typically some Web sites)
  - Need: capturing, understanding, digesting, trusting and sharing
  - e.g. report of faults written by car garages
- External information (the Web)
  - Need: capturing, understanding, contextualising, digesting, trusting and sharing
  - e.g. Information in Web pages
  - e.g. pictures provided by citizens in an emergency scenario



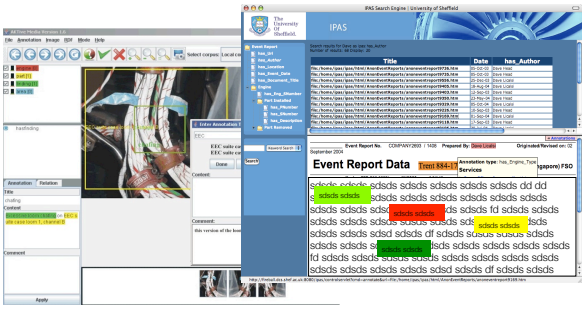
Requirements for Knowledge Acquisition

- issues in knowledge acquisition:
  - acquiring: what and what for?

The University Of Sheffield.

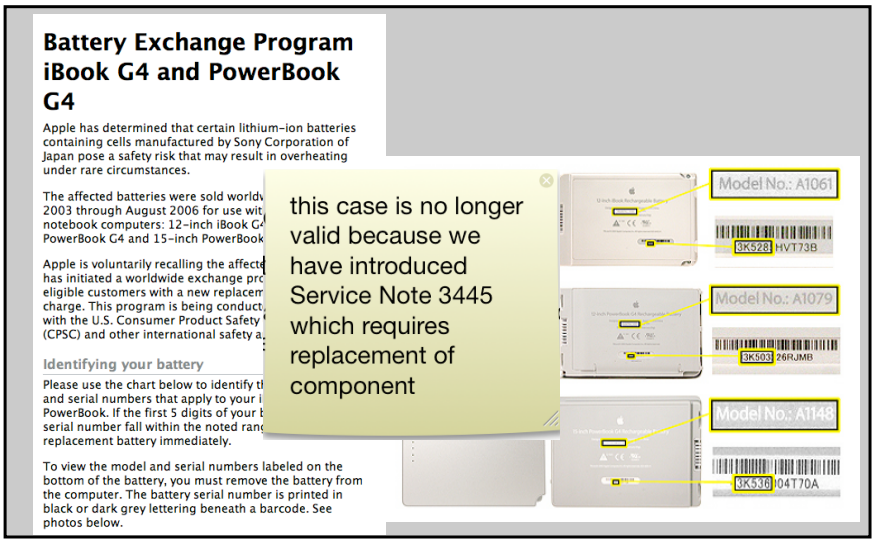
Knowledge Acquisition

- Collecting and aggregating multimedia knowledge to make it available for
  - sharing and reuse
    - From document management to knowledge management
  - for integration
- Approaches
  - At source: helping people capturing knowledge when produced
- On legacy documents, pictures, data:
  - Annotation services



Requirements for KA: Cross media

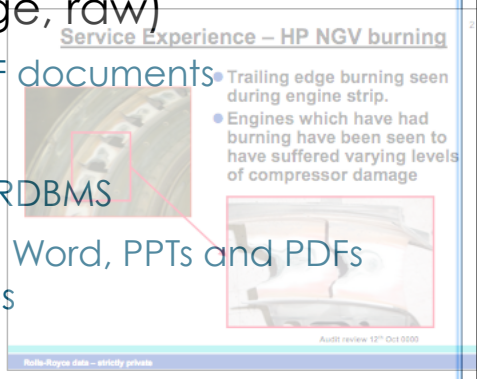
- Evidence is often distributed in different media;
- Knowledge in one medium does not carry the full evidence

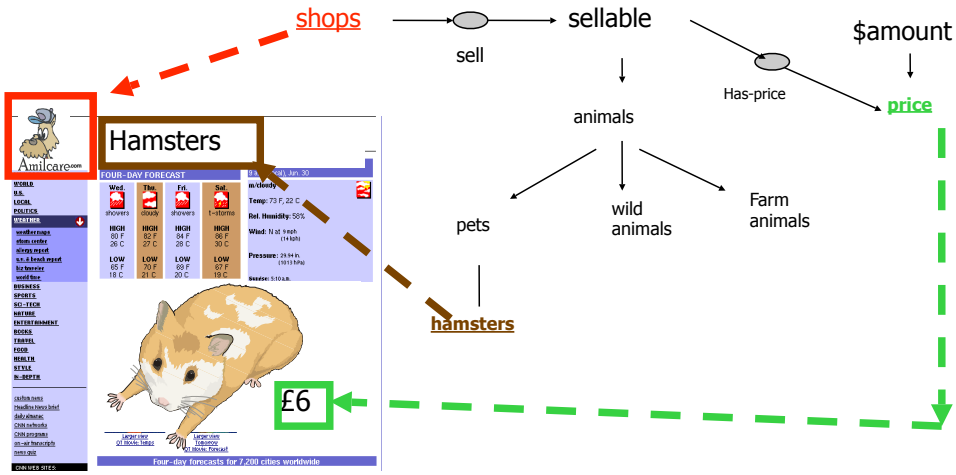


Compound Documents & CM

From Deliverable D8.2

- Typical data objects (text, image, raw)
  - Text formats: Word, Excel, PPT and PDF documents
  - Images: Jpeg and Gif
  - Raw data: Measurements stored in a RDBMS
  - Cross-media: Compound documents: Word, PPTs and PDFs containing both text and Jpeg images
    - Portions semantically related to each other within the same physical document
    - Information contained in just one modality is insufficient
    - Cross-media knowledge acquisition techniques needed in order to capture and manage all of the explicit and implicit knowledge





SW for Knowledge Acquisition

- user centred methodologies and tools for text and image annotation
- automatic methodologies and tools for text annotation

The University Of Sheffield.

Semantic Web for Knowledge Acquisition

- Aims:
  - To acquire knowledge within and across media in a rich, semantically-oriented way
  - Outcome of acquisition technologies is a semantic representation of the content (conceptualisation) to be used for knowledge management purposes
  - Enrichment of multimedia documents with layers of manually or automatically generated annotation is the main medium of associating conceptualisations to resources

The University Of Sheffield.

Annotating Resources

- 3 main methods of annotating:
  - Ontology-based annotations
  - Free text annotations - Braindumps
  - Document enrichment

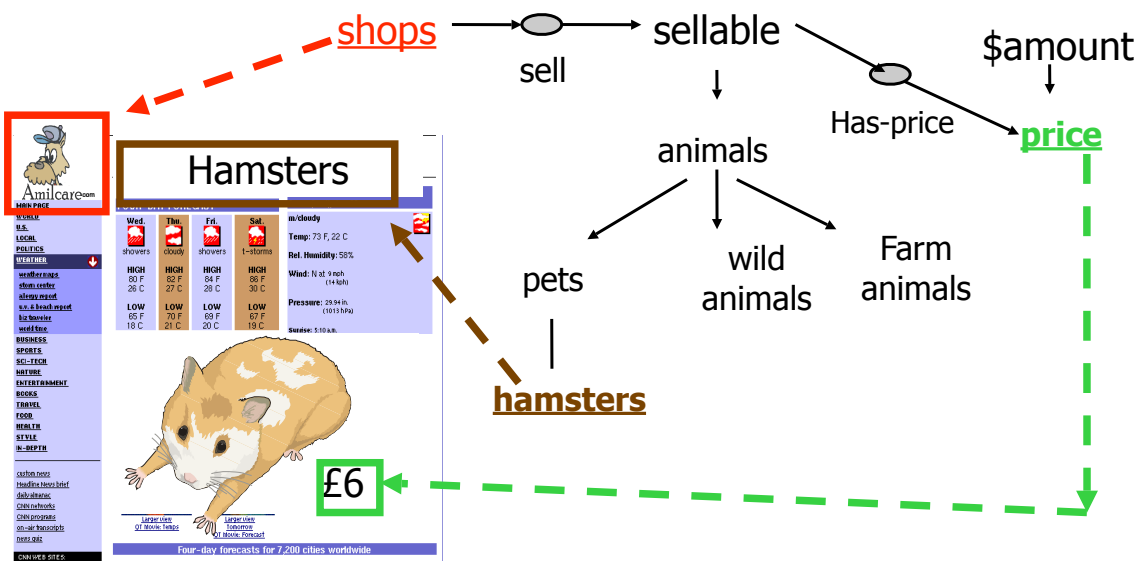
Vitaveska Lanfranchi, Fabio Ciravegna and Daniela Petrelli: Semantic Web-based Document: Editing and Browsing in AktiveDoc, 2nd European Semantic Web Conference, Crete, June 2005

The University Of Sheffield.

Ontology-based annotations

- Marking up contained information
  - Portions of documents associated to objects in ontology
    - Allows:
      - Ontology-driven processing
      - Services based on ontology will be able to use information
    - Ontomat/CREAM (Staab et al 2001)
    - Melita (Ciravegna *et al.* 2002)
    - SemTag and Seeker (Dill et al. 2003)
    - ...and many others...

Ontology-based Annotation



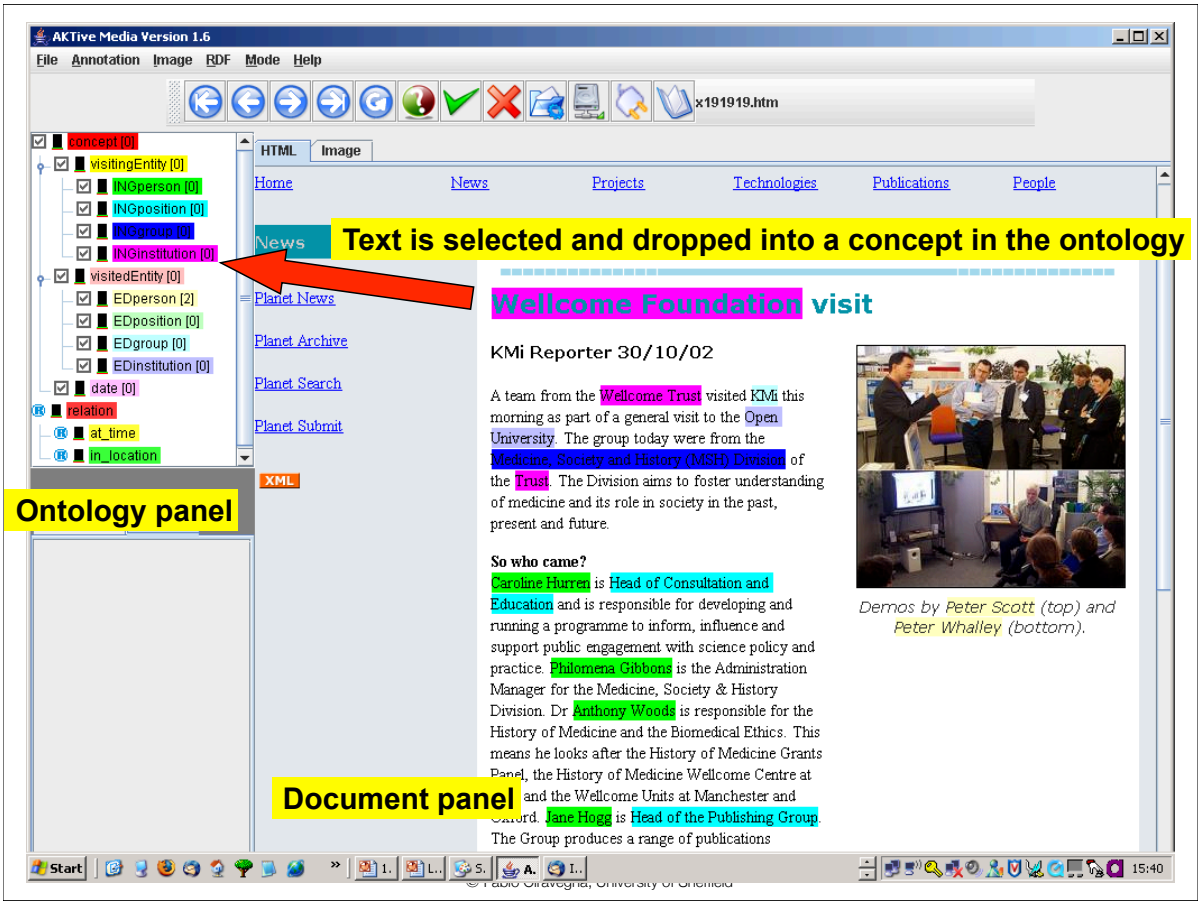
Input &Output

- Input to the KA technologies
  - Ontologies (MMO, domain ontology),
  - Background knowledge (gazetteers, etc.)
  - Normalised document representation
  - Medium to extract for (text, images, data, videos,...)
- Output
  - Evidence represented in terms of conceptual information
    - Evidence used by other modules as background conceptual knowledge, i.e. pre-existing knowledge
    - Evidence in the form of uncertain output

AktiveMedia

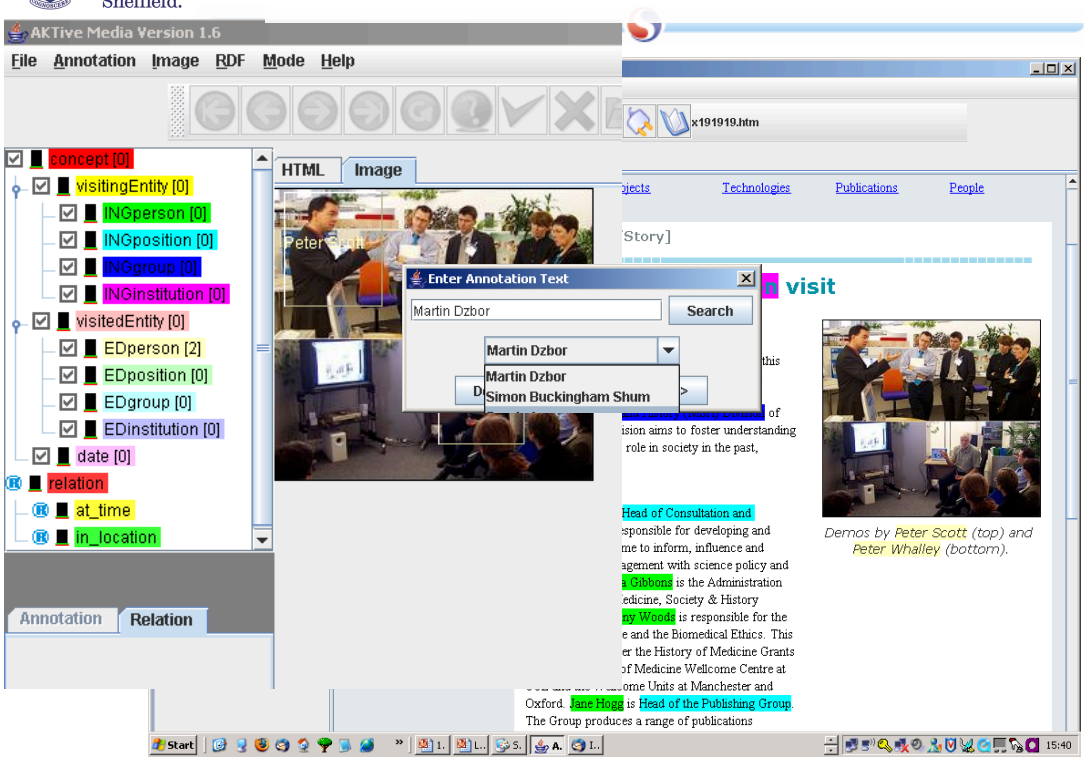
- Enables semi-automatic annotation across texts and images
- The interface enables
  - HTML editing
  - Annotation of documents in RDF based on an OWL ontology
- Types of annotations
  - Concepts / Relations
- SW: Annotation:
  - Selection of concept/relation and highlighting of text is the way in which annotation is performed

http://www.dcs.shef.ac.uk/~ajay/html/cresearch.html



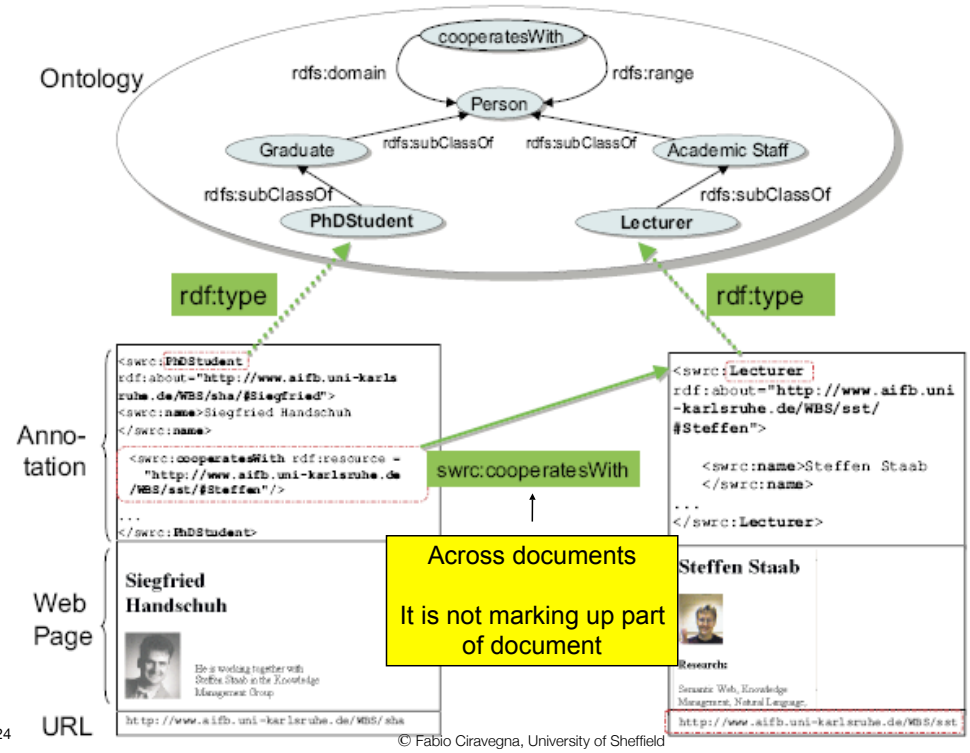
The University Of Sheffield.

Contextual Annotation of Images and Text



The University Of Sheffield.

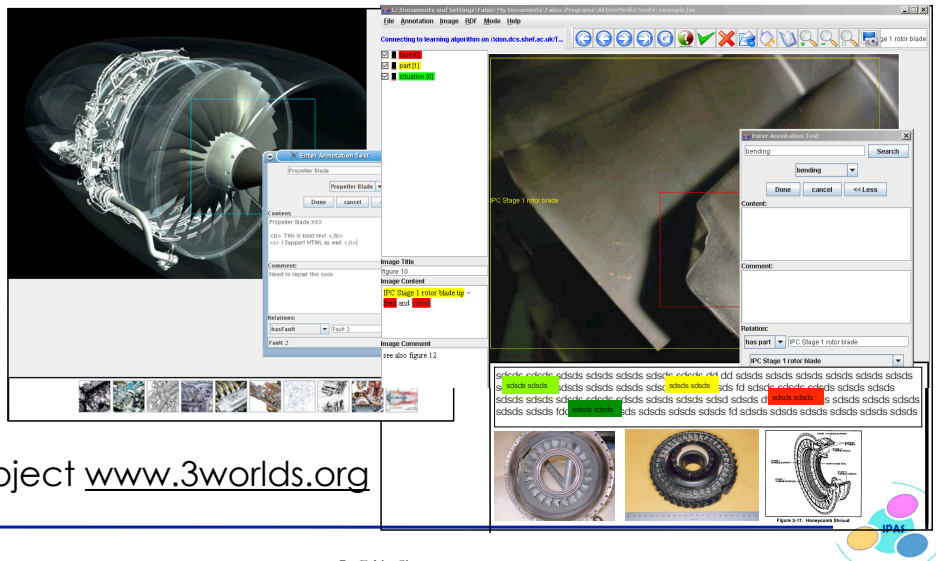
Annotating across documents (CREAM, 2001)



The University Of Sheffield.

Example of Application

- Annotation of compound documents for documenting the overhaul of a jet engine



IPAS project www.3worlds.org

The University Of Sheffield.

Annotations: Where From?

- SW relies on document annotation
  - Current state of art often requires manual annotation
- Manual Annotation
  - Very few people will annotate web pages by hand
  - What if they did?
    - Isn't the web based on hype?
      - Do people really need to publish their girlfriend photos?

The University Of Sheffield.

Manual Annotation (1)

- Expensive/time consuming/difficult
  - Chicken-egg problem
  - If it adds time to page editing, users will not do it unless there is really something for them
    - Usefulness and hype
- Inefficient and never ending
  - Every new document needs to be annotated
- Difficult
  - if two people annotate the same documents have 15-30/100 probabilities to annotate them differently
    - Risk is that the same information is annotated differently
      - Disagreement between annotators means data sparsity
      - Information becomes difficult to retrieve

The University Of Sheffield.

Problems with Manual Annotation (2)

- Tedious & Tiring
  - Error prone
- Legacy with the past
  - Ontologies are living objects, new version produced
    - Which version of the ontology is used for annotation?
- Dispersed information
  - Annotation largely unfeasible for large diverse repositories
    - E.g. a Web site
      - Department of CS of the University of Southampton: 1,600 pages
    - How many relevant ontologies are there for that department?

The University Of Sheffield.

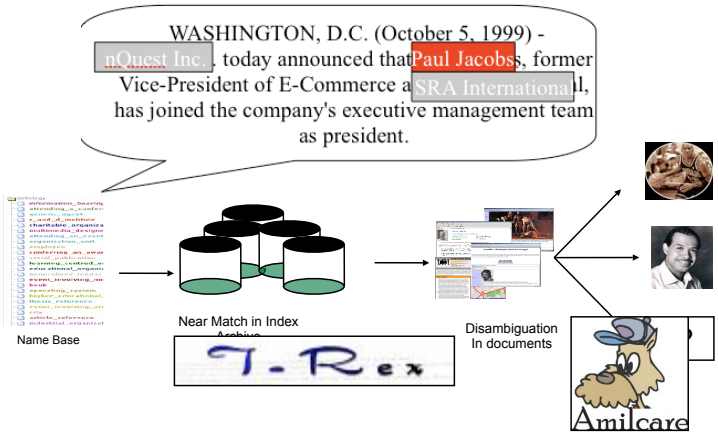
Problems with Manual Annotation (3)

- How many annotation schemas?
  - The Semantic Web is expected to be composed of
    - [Many] small ontological components [Hendler 2001] will be created, mainly related to different domain and applications
    - University of Sheffield web site:
      - What ontology for annotation?
        - Universities/Education, Research life, Scientific Papers,
        - Sport, computer network organization….
        - You name what…

The University Of Sheffield.

Annotation for use…

- If annotation is to be chosen by author/owner
  - Selection of Annotation Schema may reflect world model of the creator, not of the user
    - E.g. education is the main goal of the university, so the central Uni will probably choose an ontology on Education
    - Most of my time is actually devoted to research
    - Most of my colleagues look for scientific information on our web site
    - To us, Uni's annotation would be largely unuseful
    - Question:
      - Who (and how!) is going to introduce the annotation for us?
      - Where is the annotation to be inserted?



Automating Annotation

The University Of Sheffield.

Tasks for KA: Extraction

- Automatic annotation
  - To help manual annotation OR to replace human annotators
    - (e.g. on legacy data)
- Text:
  - Entity Extraction
  - Table Fields Extraction
  - Relation Extraction
  - Event Extraction
- Data:
  - Similarity of Data Instances
  - Functions and relation
  - Finding patterns and (ir-)regularities in data
- Images:
  - Semantically driven Image analysis using ontologies, for retrieval and annotation
  - Image classification/clustering with respect to the dominant visual trends

The University Of Sheffield.

Entity Recognition

- Tasks:
  - Recognition and classification of entities, e.g. references to concepts in document
    - E.g. people's names, companies, locations, etc.
  - Unique identification of instances (URI assignment)
    - Including disambiguation
      - Michael Jordan as basketball player Vs lawyer
      - London UK Vs London USA
  - Integration with other sources
    - E.g. positioning on a map
- This step is generally called Named Entity Recognition

The University Of Sheffield.

Traditional approach to NER

- Two steps:
  - Training phase
    - Input: annotated set of representative documents
    - Output: trained system
  - At runtime
    - One-by-one document analysis
- Expected accuracy:
  - 80-95% (free texts)
  - Web documents tend to require additional processing to get equivalent results (but do-able to some extent)
- Medium Scale: up to hundreds of thousands of documents

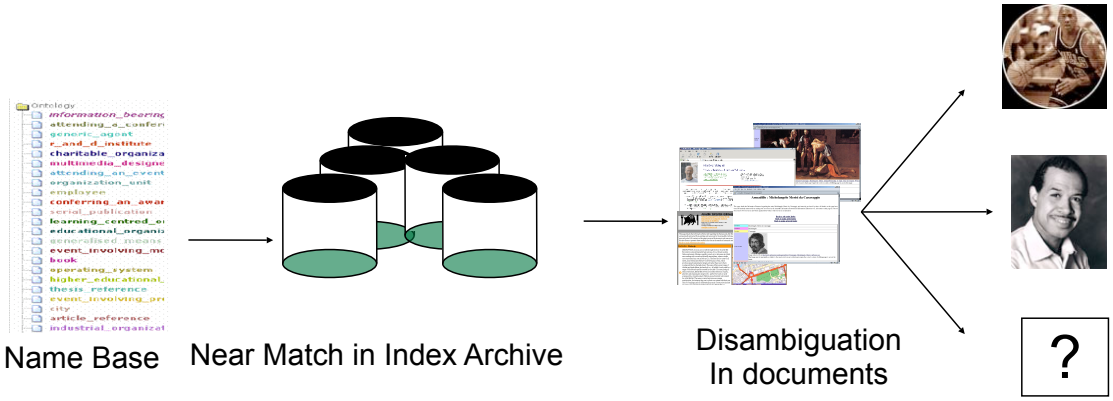
Large Scale NER

- For large scale (some hundred millions pages) smarter infrastructure is needed
  - Search engine-like indexing infrastructure
  - Faster processing (less processing)
  - Two cases:
    - Recognition of known terms (and their variations)
      - See also information integration
    - Discovery of new names

Large Scale NER: Indexing

- Document Indexing as in Search Engines

Known Name Recognition



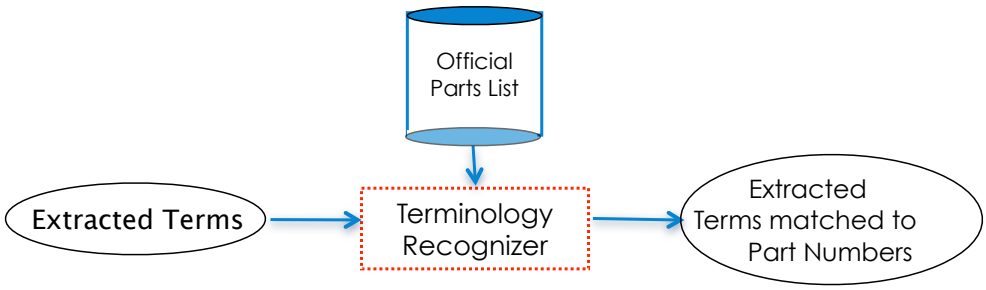
S. Dill, N. Eiron, et al: SemTag and Seeker: Bootstrapping the semantic web via automated semantic annotation. WWW'03

Discovery of New Names

- Modified Indexing of documents to recognise potential names
  - Traditional NER
    - On the window of words (not the whole doc!!!)
      - Fast and effective
  - Web specific strategies
    - To identify names without context

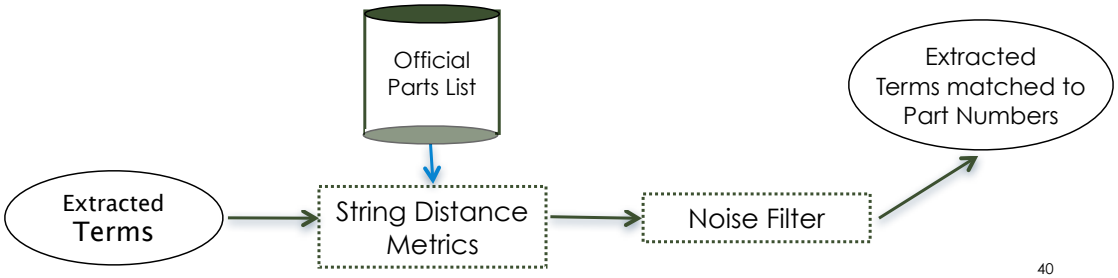
Terminology Recognition

- NER is one example of term recognition
- More useful in technical domains is terminology recognition
  - The task of assigning a URI to a technical description
    - i.e. mapping a natural language description to the official company ontology



Terminology Recognition

- Possible approaches
  - Linguistic approaches
    - Based on linguistic analysis of terms (Gaizauskas *et al* 2003)
  - Statistical approaches
    - Based on frequency analysis and detection
  - Other approaches
    - Distance metrics based (Butters 2007)



The University Of Sheffield.

More complex IE: event modelling

- Not just NER but also relation among elements in a document
  - More complex task
  - Requires some reasoning to bridge the complexity of events to the ontology structure
    - Imprecision in extraction
    - Information non matching the ontology schema
- This is where IE has hit a performance ceiling
  - 60/70 Precision/Recall ratio since 1998

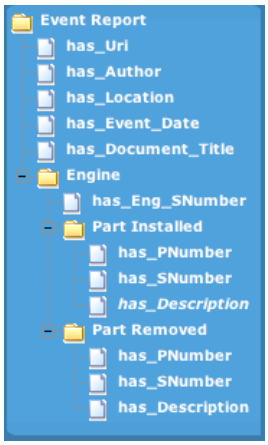
The University Of Sheffield.

Table Field Extraction

- Tables are an essential part of many documents
  - Most information is represented in tables
- Tables can be represented as forms to fill
  - Semantics is fixed
  - Wrapper writing or wrapper induction (Kushmerick 1997)
- Tables can be created ad hoc in documents (e.g. Word docs)
  - Semantics is unclear
  - Sometimes documents are created as part of a workflow, therefore they tend to be created using common models
    - e.g. by re-using the previously generated document
    - hence tables evolve, but still semantics can be traced

The University Of Sheffield.

An Example of Automatic IE



- Automatic extraction of information from event report
  - 18,000 documents analysed
- Metadata generated according to a simple ontology
- Automatic extraction of metadata and indexing of documents

http://www.3worlds.org/

The University Of Sheffield.

Types of tables in Event Reports

module/accessory details			
item	part number	s/n removed	s/n installed
	p39-401revf	04-0721257 tsn/csn: 268/106	04-1012229 tsn/csn:0/0

Part numbers
04-0721257 tsn/csn: 268/106 off
04-1012229 tsn/csn:0/0 on

s/n removed	04-0721257 tsn/csn: 268/106
s/n installed	04-1012229 tsn/csn:0/0

Parts/Components Removed or Installed (If Any):

On/Off	Part Number / Serial Number	Part Description	Hours / Cycles	Qty	Destiny Disposit
Installed	FK30840 RGG12340	TO SB72-C629)	11129 TSN 1954	1	
Installed	FK21221 EC092		11652 TSN 2119	1	
Installed	FK30840 RGG12501		11129 TSN 1954	1	
Installed	FK30840 RGG12208		11129 TSN 1954	1	
Installed	FK30840 RGG12391		11129 TSN 1954	1	

The University Of Sheffield.

Applying information extraction

- AktiveMedia to annotate texts
- TRex system (Jiria et al. 2006) to train and extract
  - http://tyne.shef.ac.uk/t-rex/
- IE captures most of the information in tables
  - 99% of the information captured (recall=99)
  - 98% of proposed information is correct (precision=98)

	POS	ACT	CORR	WRONG	MISSED	PREC	REC	F1
airport	120	120	120	0	0	100	100	100
has_airframe_cycles	104	104	104	0	0	100	100	100
has_airframe_hours	104	104	104	0	0	100	100	100
has_author	120	120	120	0	0	100	100	100
has_engine_serial_number	120	120	120	0	0	100	100	100
has_engine_type	120	120	120	0	0	100	100	100
has_event_date	120	120	120	0	0	100	100	100
has_event_report_no	356	358	356	2	0	99	100	100
has_part_description_installed	120	113	111	2	9	98	93	95
has_part_description_removed	120	133	120	13	0	90	100	95
has_part_number_installed	120	113	111	2	9	98	93	95
has_part_number_removed	120	133	119	14	1	89	99	94
TOTAL	**1644**	**1658**	**1625**	**33**	**19**	**98**	**99**	**98**

Information Integration

The University Of Sheffield.

Information Integration

- Facts from different sources need to be integrated
  - To connect information/knowledge across docs
    - Assign unique URI
  - To solve discrepancies and ambiguities
- Steps
  - Unique instance identification (for entities)
  - Record linkage (for events)
- Information Integration strategies
  - Generic
    - Distance metrics (Chapman 2004)
    - Using Web bias
  - Statistical matching
  - Application specific
    - Rules

The University Of Sheffield.

Armadillo: Historical Data Mining

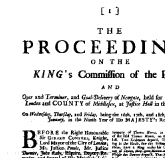
Arts & Humanities Research Council

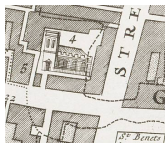
AHDS Deposits

- The Marine Society Registers
- The Westminster Historical Database
- Eighteenth Century Fire Insurance Policies
- Prerogative Court of Canterbury Wills
- The Proceedings of the Old Bailey

- St. Martin's Settlement Exams Index — WESTCAT
- Collage image databse — Guildhall Library
- Harben's Dictionary of London
- John Strype's "Survey…"
- Metropolitan London in the 1690s — IHR
- Selected Criminal Records — PRO
- House of Lords Journals — BOPCRIS

http://www.motco.com



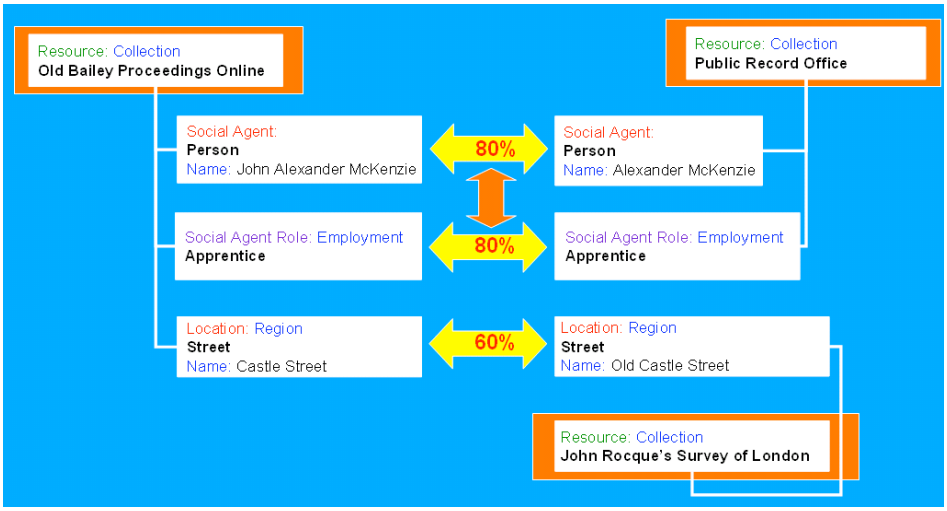


http://www.hrionline.ac.uk/armadillo/

The University Of Sheffield.

Information Integration

Armadillo: Historical Data Mining

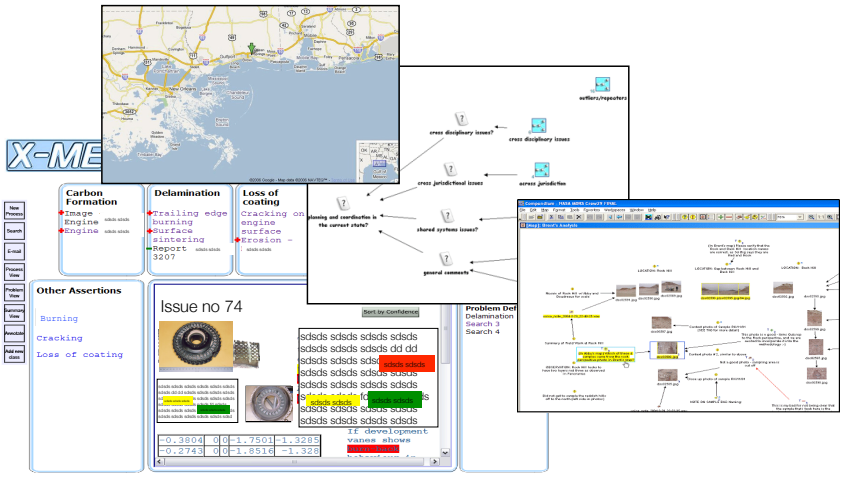


Arts & Humanities Research Council

The University Of Sheffield.

Conclusions on KA, Requirements and Tools

- Large scale?
  - Ontologies:
    - large ontologies (up to 10k) with simple tasks (SemTag and Seeker, Kim)
    - small/medium scale (up to 100) with more complex tasks
  - KB: large scale
- Portability: most technology difficult to port without experts (Armadillo, KIM)
  - User input well exploited in human-centred acquisition (e.g. Melita, AktiveMedia)
- Cross-Media: exploited in user centred annotation (e.g. AktiveMedia)
- Background Knowledge
  - Used in AktiveMedia, KIM, SemTag and Armadillo to some extent
  - Uncertainty: some use in Armadillo



Knowledge Sharing and Reuse

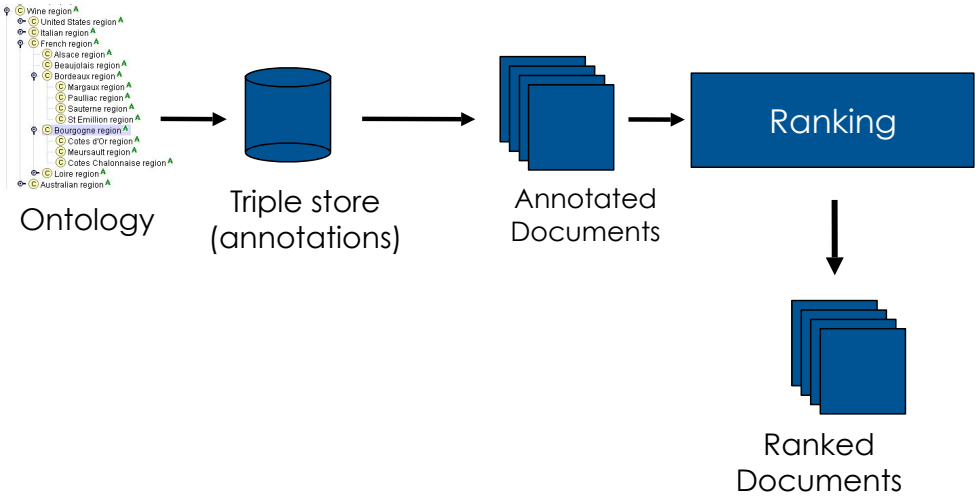
- issues in knowledge sharing
- approaches and novel methods to searching, sharing and reuse knowledge

The University Of Sheffield.

Sharing and Reuse via SW

- Ontology based annotation enables
  - Searching using ontologies
    - Searching metadata rather than text
  - Connection of information across documents, media and archives
    - Retrieving information independently from the store/media
  - Reasoning on knowledge
    - Making implicit explicit
  - Workflow support
    - Supporting user actions rather than single searches

Searching Documents using ontologies



Does it work?

- An Experiment on Jet Engine Event Reports
  - 21 topics of search, e.g.
    - "How many events were caused during maintenance in 2003?"
    - "What events were caused during maintenance in 2003 due to control units?"
    - 'Find al l the events associated with damage to acous- tic liners fol lowing bird strike"
  - How many topics can we model with Information Extraction?
    - 21 topics/ 14 topics partially or not covered by IE-based annotations
      - given size of corpus there is no way that manual annotations are added

The University Of Sheffield.

Results for ontology matching for even reports

- 85% of documents in the first 20 hits are relevant
  - Compare with keywords: 56%
- 40% of relevant documents are in the first 2 pages
  - Compare with keywords: 57%

- Ontology matching implies
  - Reading a limited amount of irrelevant documents
  - Risking missing many documents
  - It is possible to *count* the events

The University Of Sheffield.

Hybrid Search (keywords+ontology)

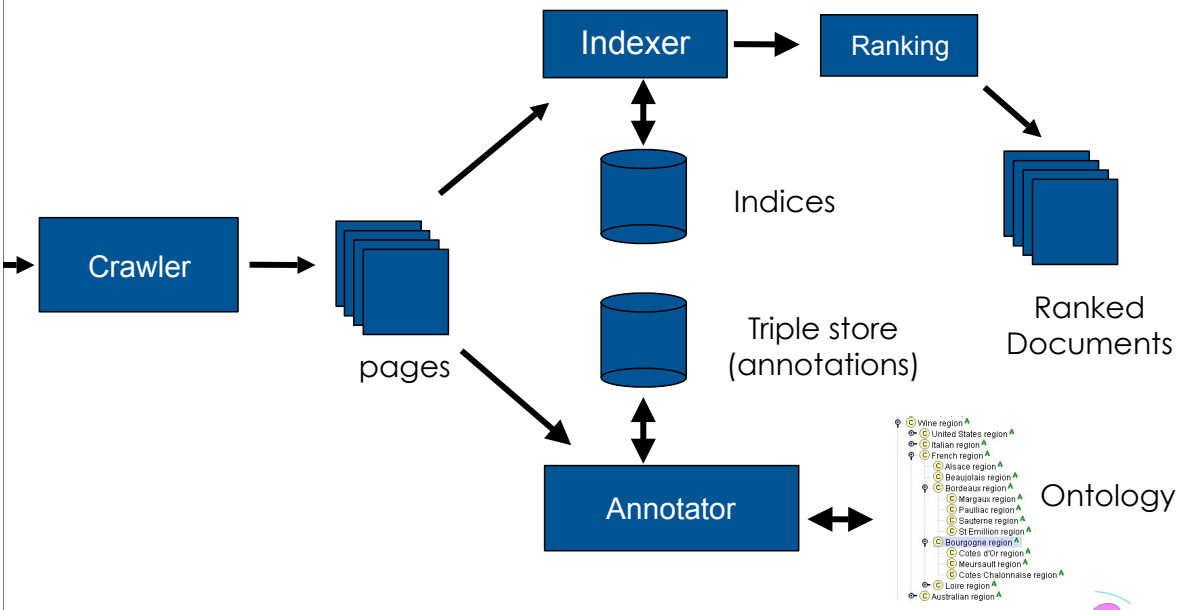
- Mixes keyword and ontology based search
- Ontology based search
- Traditional keyword search
- Keyword in contest of ontology-based annotations
- Potential queries:
  - Return all documents where the word fuel is mentioned
  - Return all documents where the affected part description includes the word fuel

    affected parts is concept in ontology
  - Return all documents where the affected part description is similar to "fuel duct"
  - Return all documents where the affected part description is equal to "fuel duct" (URI=XXXXX)

Vitaveska Lanfranchi, Ravish Bhagdev, Sam Chapman, Fabio Ciravegna, Daniela Petrelli: Extracting and Searching Knowledge for the Aerospace Industry, in Proc. of 1st European Semantic Technology Conference, Vienna, May 2007

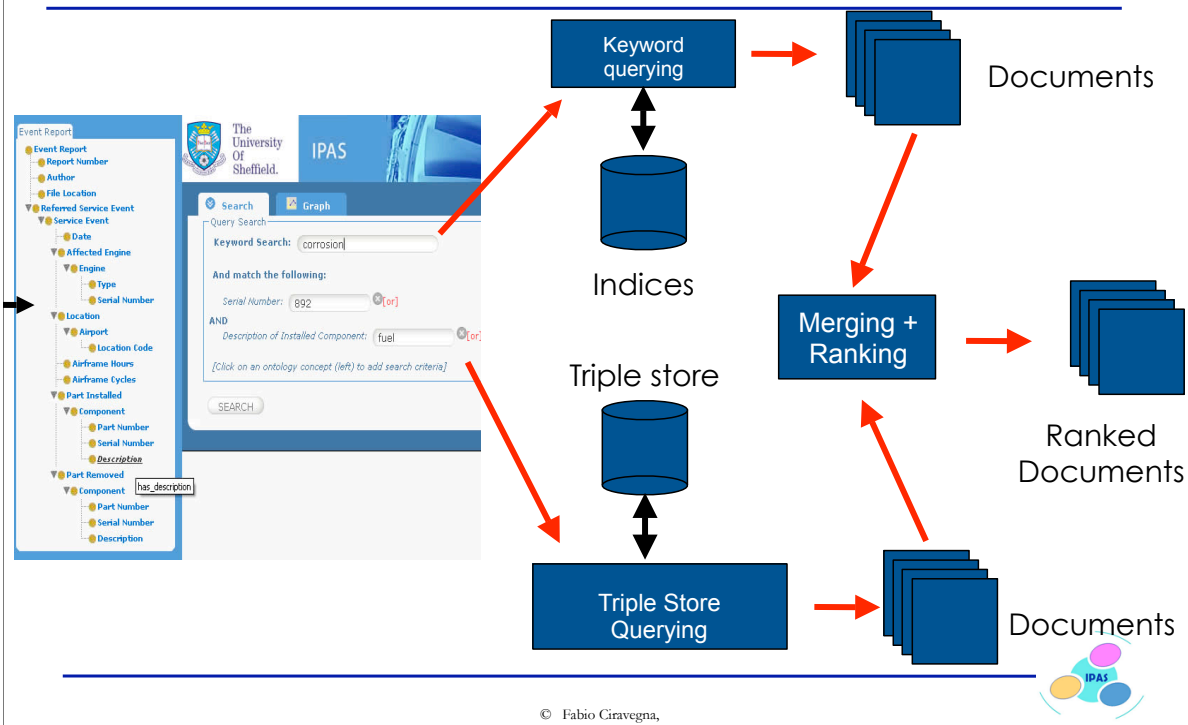
The University Of Sheffield.

Hybrid Indexing/Annotation



The University Of Sheffield.

Hybrid Search



The University Of Sheffield.

Advantages with Hybrid Search

- Accuracy of Ontology-based searching available
  - When metadata covers information
- Expressiveness of Keyword querying is available
  - For all other cases
- Keyword-in-context available
  - Keyword matching available for matching concepts names
    - e.g. match "fuel" in the description of the removed parts
  - Uses provenance of annotations
    - Portion of document annotated with concepts are stored in 3store
    - Keyword matching applied only on the relevant strings
      - e.g. "fuel" is matched only on snippets of texts annotated as removed parts

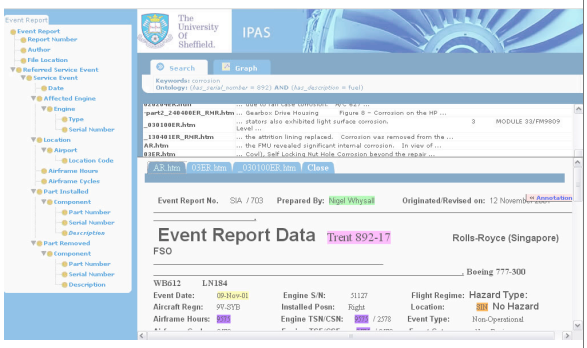
The University Of Sheffield.

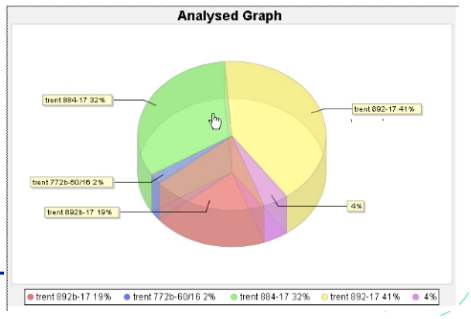
Results for Hybrid Search

- 83% of documents in the first 20 hits are relevant
  - K:56% O:85%
- 85% of relevant documents are in the first 2 pages
  - K: 57% O:47%
- F(1)=84%
  - K:57% O:54%
- Hybrid Search implies
  - Reading a limited amount of irrelevant documents
  - Being able to retrieve easily a very large part of documents

X-Search: implementation of HS

- Enables
  - Flexible access to metadata and legacy documents via Hybrid Search
    - Users can choose their own search strategy
  - Enables quantification of events via graphs
- Supported by
  - Keyword indexing
  - Automatic generation of metadata via IE (via TRex)
  - User-centred semi-automatic annotation (via aktiveMedia)
- Currently in Beta test by hundreds of Rolls Royce engineers



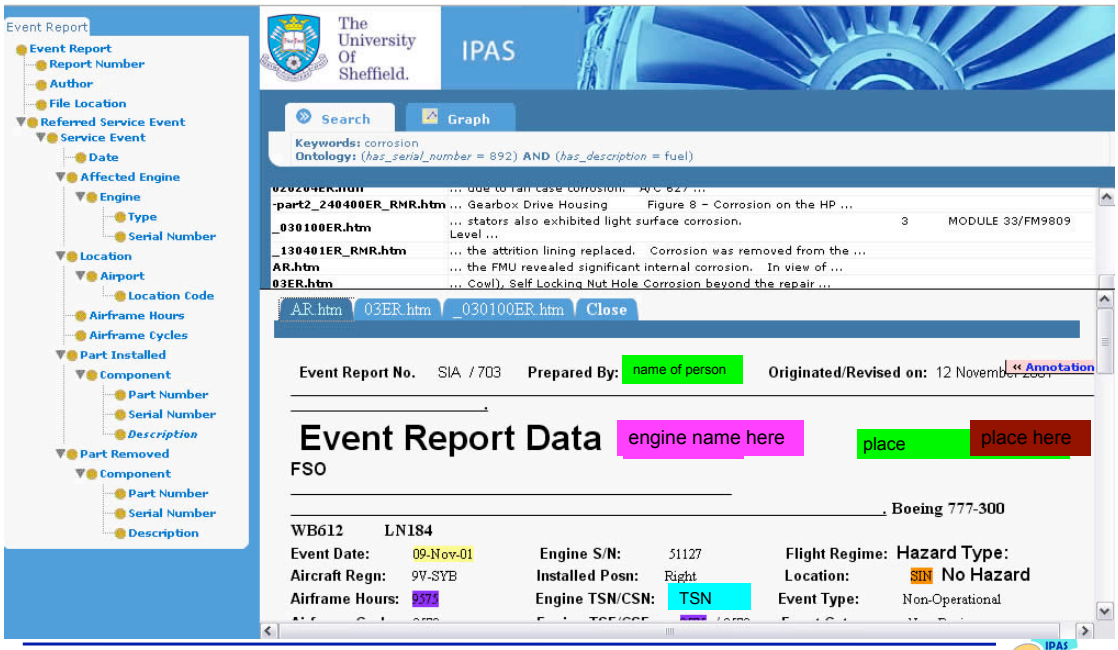


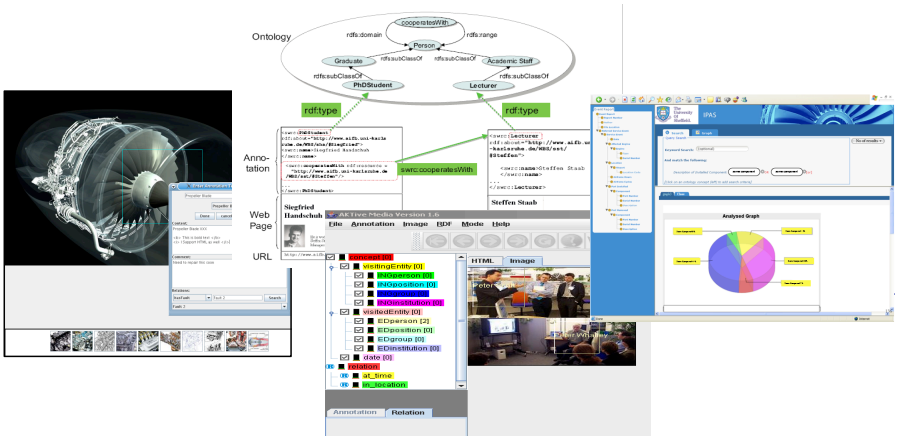
Query results

- Results are displayed as a list
- User can click on a document and open it in the lower frame
- The document will be enriched by annotations with attached services
- Multiple documents can be opened in a tab interface

The University Of Sheffield.

Query results





Conclusions

The University Of Sheffield.

Conclusions

- Document annotation can be performed at different levels
  - Ontology-based, braindump, document enrichment
- Annotation unlikely to be performed manually on a large scale except for limited cases (e.g. FoaF)
- Automation can be applied successfully for helping annotating
- We have seen:
  - User centred automated ontology-based annotation
    - For trusted self contained documents (e.g. KM)
  - Automatic document Enrichment
    - Melita/Magpie/AktiveDoc
  - Unsupervised large scale annotation
    - For distributed large scale environments (e.g. the Web)
    - SemTag&Seeker, Armadillo

The University Of Sheffield.

Future Work & Challenges

- Multidisciplinary research for automation
  - NLP has strong role, but complemented with other disciplines
  - SE, ML, II, SWS, HCI
- Annotation
  - Beyond the division between user centred and unsupervised
    - Strong HCI strategies
      - Validation of results across documents
        - How can you validate 2M triples produced by large scale annotation?
- Information extraction models
  - Beyond simple IE models
    - Towards fully fledged adaptive IE systems
      - Maintaining flexibility
- Information Integration
  - Towards complex trainable strategies for integration
- Combination of evidence
  - Of sources
  - Of extractors

The University Of Sheffield.

Future Work & Challenges (2)

- How modelling uncertainty?
- Knowledge is dynamic. How do you model that?
- HCI
  - Information presentation (document annotation)
    - Intrusivity:
      - How to avoid annoying users with too many annotations
    - Trust
      - Who do users trust?
        - Tracing preferred sources
      - Where does the information come from?
- Scalability
  - Large scale indexing systems
    - Millions of pages (not billions!)

The University Of Sheffield.

Conclusions and Future Work

- Knowledge Management is moving towards large scale
  - Initially expected around 2010 now already happening
- The Semantic WEB offers <u>potentially</u> key technologies to the development of future KM
  - More Web than Semantics, but:
    - A little semantics goes a long way (J. Hendler)
- The potential must be exploited addressing <u>real world</u> requirements
  - Rather than in principle AI-oriented requirements (e.g. closed world, small scale, etc.)
- Strong application pull can be obtained
  - Do not sell slogans, sell ideas and applications!

Rolls-Royce Kodak Environment Agency LYCOS

A final thought

- These technologies allow easy collection of *very* large amount of information/knowledge
- Are we:
  - Preparing for a better Web/better world?
  - Preparing for a world with no privacy?
    - Big brother
    - Spam
    - Identity theft
  - Just adding hay to the haystack while searching for a needle?
    - Drowning in triples while trying to avoid drowning in texts?

The Karen Spark-Jones slide

Thank You

- Contact Information
  - www.dcs.shef.ac.uk/~fabio
  - fabio@dcs.shef.ac.uk
- Intelligent Web Technologies Lab
  - http://nlp.shef.ac.uk/wig/
- NLP Sheffield
  - http://nlp.shef.ac.uk/
- University of Sheffield
  - www.shef.ac.uk

The University Of Sheffield.

A very Incomplete Bibliography

- F. Ciravegna: Challenges in Information Extraction from Text for Knowledge Management, in S. Staab, (ed), "Human Language Technologies for Knowledge Management", IEEE Intelligent Systems and Their Applications (Trends and Controversies), Vol. 16, No. 6, pp 88-90, 2001.
- Fabio Ciravegna. Adaptive information extraction from text by rule induction and generalisation. In Proceedings of 17th International Joint Conference on Artificial Intelligence (IJCAI), 2001. Seattle.
- H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02). 2002.
- I. Muslea, S. Minton, and C. Knoblock. 1998. Wrapper induction for semistructured webbased information sources. In Proceedings of the Conference on Automated Learning and Discovery (CONALD), 1998.
- Vitaveska Lanfranchi, Fabio Ciravegna, Daniela Petrelli: Semantic Web-based Document: Editing and Browsing in AktiveDoc, Proceedings of the 2nd European Semantic Web Conference , Heraklion, Greece, May 29-June 1, 2005
- Handschuh, Staab, Ciravegna. S-CREAM - Semi-automatic CREAtion of Metadata (2002) http://citeseer.nj.nec.com/529793.html
- F. Ciravegna, A. Dingli, D. Petrelli, Y. Wilks: User-System Cooperation in Document Annotation based on Information Extraction. Knowledge Engineering and Knowledge Management (Ontologies and the Semantic Web), (EKAW02), 2002.
- M. Vargas-Vera, Enrico Motta, J. Domingue, M. Lanzoni, A. Stutt, and F. Ciravegna. MnM: Ontology driven semi-automatic or automatic support for semantic markup. In Proc. of the 13th International Conference on Knowledge Engineering and Knowledge Management, EKAW02. Springer Verlag, 2002

The University Of Sheffield.

A very Incomplete Bibliography (ctd)

- Fabio Ciravegna. Designing adaptive information extraction for the Semantic Web in Amilcare. In S. Handschuh and S. Staab, editors, Annotation for the Semantic Web, Frontiers in Artificial Intelligence and Applications. IOS Press, 2003.
- C. Goble, S. Bechhofer, L. Carr, D. De Roure, and W. Hall. Conceptual Open Hypermedia = The Semantic Web? In The Second International Workshop on the Semantic Web, pages 44–50, Hong Kong, May 2001
- Fabio Ciravegna, Sam Chapman, Alexiei Dingli, and Yorick Wilks: Learning to Harvest Information for the Semantic Web, Proceedings of the First European Semantic Web Conference, Crete, May 2004
- A. Kiryakov, B. Popov, et al. Semantic Annotation, Indexing, and Retrieval. 2nd International Semantic Web Conference (ISWC2003), http://www.ontotext.com/publications/index.html#KiryakovEtAl2003
- S. Dill, N. Eiron, et al: http://www.tomkinshome.com/papers/2Web/semtag.pdf . SemTag and Seeker: Bootstrapping the semantic web via automated semantic annotation. WWW'03.
- Thomas Leonard and Hugh Glaser. Large scale acquisition and maintenance from the web without source access. In Siegfried Handschuh, Rose Dieng-Kuntz, and Steffen Staab, editors, Proceedings Workshop 4, Knowledge Markup and Semantic Annotation, K-CAP 2001, 2001
- Martin Dzbor, John B. Domingue, and Enrico Motta. Magpie - towards a semantic web browser. In Proceedings of the 2nd Intl. Semantic Web Conference, October 2003. Sanibel Island, Forida
- Alexander Maedche, Steffen Staab, Nenad Stojanovic, Rudi Studer, York Sure: SEmantic portAL - The SEAL approach In D. Fensel, J. Hendler, H. Lieberman, W. Wahlster (eds.), Spinning the Semantic Web, pp. 317-359. MIT Press, Cambridge, MA., 2003.

The University Of Sheffield.

A very Incomplete Bibliography (ctd)

- Natalya F. Noy and Deborah L. McGuinness: Ontology Development 101: A Guide to Creating Your First Ontology, http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html
- Elena Paslaru Bontas, Christoph Tempich, York Sure : OntoCom: A Cost Estimation Model for Ontology Engineering, In: Proceedings of the 5th International Semantic Web Conference (ISWC 2006), November 5-9, 2006, Athens, GA, USA, LNCS. Springer.
- Ajay Chakravarthy, Vita Lanfranchi and Fabio Ciravegna: Cross-media Document Annotation and Enrichment, SAAW2006 - 1st Semantic Authoring and Annotation Workshop, The 5th International Semantic Web Conference (ISWC2006), Athens, GA, USA, Monday, November 6th 200
- R. Gaizauskas and G. Demetriou and P. Artymiuk and P. Willett: Protein Structures and Information Extraction from Biological Texts: The PASTA System, Journal of Bioinformatics 19(1), 135-143, 2003
- Vitaveska Lanfranchi, Ravish Bhagdev, Sam Chapman, Fabio Ciravegna, Daniela Petrelli: Extracting and Searching Knowledge for the Aerospace Industry, in Proc. of 1st European Semantic Technology Conference, Vienna, May 2007