

BIOMEDICAL NLP WORKSHOP AT RANLP 2017



ENTITY-CENTRIC INFORMATION ACCESS WITH THE HUMAN-IN-THE-LOOP FOR THE BIOMEDICAL DOMAINS

SEID MUHIE YIMAM, STEFFEN REMUS, ALEXANDER PANCHENKO, ANDREAS HOLZINGER, AND CHRIS BIEMANN

SEPTEMBER 8, 2017



08.09.2017 Language Technology Group

Outline



- Introduction
- Entity centric information access
- Case studies
 - Adaptive annotation WebAnno
 - Collection insights New/s/leak
 - Personalized Knowledge management Storyfinder
- Human-in-the-Loop information management
- Conclusion



Introduction - 1



- Issues with knowledge management (KM):
 - Knowledge bottleneck (resources not available)
 - Large burden on the human
 - Modeling of entire domains





- Medical researchers process enormous amount of literatures - PubMed adds about half a million papers each year
- We propose to reload the traditional top-down KM approach and replace with a problem-oriented bottom-up approach – for biomedical domains

Introduction - 3

- Objective



- Entity Graph: bottom-up construction of a problem-oriented ontology
 - Help to document relations between biomedical entities
 - Generate new relations (semi-automatic)
 - Interactive and graph based communications
 - Share knowledge and annotations among researchers



Entity Graph Example



Universität Hamburg

6 📕

Entity centric Information Access









- WEBANNO
- New/s/leak
- StoryFinder





WEBANNO

- New/s/leak
- StoryFinder





- Adaptive Annotation – WebAnno 1/1



 WebAnno is a general purpose web-based annotation, automation, correction and curation tool for a wide range of annotations needs.





- Adaptive Annotation – WebAnno 1/2



- Focused dataset suiting the domain
- Annotate ->predict -> verify
 - Start annotation (without pre-annotation A)
 - Use annotations for automation and provide suggestions – B
 - Annotators accept correct annotations and add new ones - C







- Entity annotation
 - focused on B-Chronic lymphocytic leukemia (B-CLL)
 - Expert selects relevant documents
 - Entity types determined during annotation





- Adaptive Annotation WebAnno 1/4
- Suggestions after annotating 5 abstracts



(b) Automatic suggestions after 5 abstracts are annotated.





- Adaptive Annotation – WebAnno 1/5

Suggestions after annotating 9 abstracts







- Adaptive Annotation – WebAnno 1/6



Relation annotation – cause and effect of disease



I Long-term inositol phosphate release, but not tyrosine kinase activity, correlates with IL-2 secretion and NF-AT induction in





- WEBANNO
- New/s/leak
- StoryFinder



- New/s/leak – (NetWork of Searchable Leaks) 2/1





08.09.2017

18

- New/s/leak 2/2



- A tool designed to support investigative data journalism by exploring large sets of typically leaked documents
- Named entities are automatically identified and ranked by importance.
- A global graph of entities is used to display high-level interactions among those entities
- Set of interactions: full text search, entity merger or removal, document aggregation using meta-data ...

- New/s/leak 2/3



- The graph view: shows named entities and their relations
- The document timeline view: shows document frequency in different epochs
- The document view: is composed of the document list and a document text for reading
- The metadata view: which offers different metadata for filtering relevant or irrelevant documents.

- New/s/leak 2/4



New/s/leak UI components based on the ENRON dataset







- WEBANNO
- New/s/leak
- StoryFinder



- Storyfinder 3/1





23

- Storyfinder 3/2



- Objectives:
 - Browse: Which information is contained on a website?
 - <u>Recover</u>: Which website(s) contained the information?
 - <u>Context</u>: What else do we know about the entities?
- Originally targeted for news texts

* **Storyfinder**: Steffen Remus, Manuel Kaufmann, Kathrin Ballweg, Tatiana von Landsberger, and Chris Biemann (2017) Personalized Knowledge Base Construction and Management by Browsing the Web, in CIKM 2017, Singapore (to appear)

- Storyfinder – Concepts 3/3



- Browser plugin
 - User oriented
- Extract information while visiting websites
 - Named Entities
 - Keywords
 - Relations between entities
- Visualization
 - Node-link graph in sidebar
 - Highlighting on website
- Editable graph





- Storyfinder - Schema 3/4



- The Mozilla Firefox browser plugin:
 - Listens and reacts to a user's actions
 - Initiates the analysis of a currently visited webpage on the backend server
 - provides a side pane view to visualize the collected information
- The server backend: analyze webpage, extract and store metadata
- The interactive web page: provides real-time access to information



- Storyfinder – Functionalities 3/5





Human-in-the-Loop Information Management



- Use entity graph representation as the primary means for visualizing and accessing biomedical research documents.
- Put the user in the center of the process and integrate adaptive ML approach
- The biomedical researcher can utilize the entity graph as a visually supportive notepad.
- Automatic methods only provide suggestions, experts are in full control of the entity graph
- Entity graphs can be shared among researchers

Human-in-the-Loop Information Management[3]



• An entity graph summarizing the literature research on B-CLL



Human-in-the-Loop Information Management[4]



- Bottom-up approaches to construct a global entity graph of a domain in an incremental and problem-driven way.
- Provide provenance information through edges of global EGs (i.e. who has entered information, based on which document) for mutual understanding.
- The global graph can also incorporate information from resources, such as MeSH, Gene and Disease Ontology

Human-in-the-Loop Information Management[5]



- Challenges in the adaptation include:
 - A high quality tagging of biomedical entities
 - Preprocessing
 - Design of the user interface
 - Responsive online-adaptive machine learning model.

Conclusion



- Entity-centric information extraction and access for biomedical documents
- WebAnno: interactive and adaptive entity extraction
- New/s/leak: interactive navigation for investigative entity exploration
- Storyfinder: personalized information and knowledge management
- All can be combined, adapted, and re-focused to provide a data driven, bottom-up, conceptualization
- Human-in-the-loop is an integral component



Thank you! Question?



08.09.2017