

**Appropriating SHOT's classification in
the context of the construction of the
'history ontology' of a second
generation historical informatics
research tool:
Examples from research on computing
and biotechnology**

Akrivi Katifori

Nadzeya Kiyavitskaya

Costas Morfakis

Giannis Binietoglou

Outline

- ▶ Motivation
- ▶ Papyrus approach
- ▶ Two ontologies
 - ▶ History ontology
 - ▶ News ontology
 - ▶ Bridging History and News
- ▶ Browsing and searching history with Papyrus
- ▶ Conclusions and future work

Motivation

- ▶ Massive availability of multimedia archival material in digital form
- ▶ Need for advanced search and retrieval
- ▶ “Second generation” digital information retrieval tools aim to provide an effective way to access archival material for historians

Papyrus approach

- ▶ Papyrus is an interdisciplinary research project that investigates the use of media for historical research and proposes solutions to support this research

www.ict-papyrus.eu

- ▶ Case studies are based on archival news material from
 - ▶ AFP (Agence France Press)
 - ▶ Deutsche Welle
- ▶ Two domains of focus
 - ▶ Biotechnology and biomedical technology
 - ▶ Renewable energy with focus on wind power
- ▶ Three languages
 - ▶ English, German, French

Papyrus approach:

Study of research method

Example: *“I am interested in information on the changes in biotechnology from the beginning of the 20th century until 1970”*, to find the answer one need to:

- ▶ **Collect relevant secondary material**
 - ▶ essays of other history researchers on related subjects
 - ▶ common vocabularies containing historiographical issues e.g. “Controversies and Disputes” or “Discipline formation”
- ▶ **Collect primary material**
 - ▶ news archive content related
 - ▶ vocabulary about the news creation

Problem: historians and news professionals use different languages !

Papyrus approach:

Concepts in time and space

▶ Evolution

- ▶ Concepts change their meaning with the passage of time

Example: “Biotechnology” was related to food technology and fermentation in the beginning of 20th century, whereas now its meaning has progressed to genetics and biomedical engineering

[Robert Bud “Biotechnology in the Twentieth Century”, Social studies of science, 1991, Vol. 21, p. 415-456]

▶ Multilingualism

- ▶ One term is introduced in different times in different countries

Example: “biotechnology” has undergone different development paths in German-speaking and English-speaking countries

- ▶ Different term evolution, i.e., during the same time period a term could mean different things in different languages

Two ontologies

- ▶ An Ontology for Informatics is a formal model of a domain of interest
 - ▶ Relevant concepts are put into a categorization
 - ▶ Concepts are linked through relationships among them
 - ▶ Concepts have properties
- ▶ Ontologies offer a level of formalism that is both readable by humans and computers
- ▶ We used ontologies as tools to model our domains:
History and News

History ontology

- ▶ We referred to two history societies
 - ▶ History of Science Society and its journal *ISIS*
 - ▶ Society for the History of Technology and its journal *Technology and Culture*
- ▶ They were used as sources to provide
 - ▶ Periodization: a list of time periods and chronological divisions was obtained
 - ▶ Classification: subject index was used to find relevant concepts
- ▶ CIDOC Conceptual Reference Model adopted as a basic formalism to structure identified concepts

History ontology: general subjects of issues

1. **change in science/technology**: change in science, change in technology, environmental history, discipline formation, discovery (in science), artifacts, experiments and experimentation, academic disciplines, scientific communities, professions and professionalization

2. **institutions**, universities and colleges, societies, institutions, academies, (international) congresses, conferences, and meetings, research institutes, research schools, research stations, laboratories, prizes, awards, Nobel Prizes

3. **research and development**, technological innovation, impact of technology, technology assessment, public policy, government sponsored science, patents, big science, science and industry, technology and industry, entrepreneurs and entrepreneurship

History ontology: general subjects of issues

4. **controversies and disputes**, determinism, progress (ideas of), revolutions in science, globalization, modernization, international cooperation, futurism, utopias, authority of science, technocracy, controversies and disputes, political activists, non-governmental organizations, risk assessment, biological diversity, safety, limits of science

5. **popularization**, popular culture, rhetoric, metaphors and analogies, public opinion, public understanding of science, expert testimony

6. **ethics**, science and ethics, technology and ethics, privacy, private life, interprofessional relations

News ontology

- ▶ The News ontology is a tool to be used for news item categorization and retrieval by the News Agencies
- ▶ Main structures
 - ▶ NewsML-G2 format constructs developed for storage and retrieval of news content by International Press Telecommunications Council (IPTC)
 - ▶ Semantic concepts carried by the news items from two specific domains
 - ▶ Biotechnology and biomedical technology
 - ▶ Renewable energy with focus on wind power

Bridging History and News

- ▶ Two domains are linked through mappings, i.e., correspondences between concepts of the News and History ontologies
- ▶ Automated mapping retrieval is difficult
 - ▶ Example: “public opinion on stem cells” in the History ontology vs. “public opinion” and “stem cells” in the News ontology
- ▶ Mappings are established manually via TrenDS tools

Extended Papyrus Browser View

Information need: "Public opinion on stem-cells"
Selection: Public Opinion, Stem-cell

The screenshot displays the Extended Papyrus Browser interface with several panels:

- Find news:** Includes a search bar and a 'Search' button.
- Domains:** A list of domains with checkboxes. 'Biotechnology' is selected.
- Historiographical Issues:** A tree view of issues. 'Public opinion' is selected.
- History Ontology:** A list of related history concepts. 'Stem cell' is selected.
- History Properties:** A table showing properties for the selected concept. A tooltip is visible for the 'has definition' property.
- News Ontology:** A list of related news concepts. 'Stem cells' is selected.
- News Items:** A table of news items related to the selected concept.

History Properties Table:

Property	Value
has definition	Stem cell is a cell from which

News Items Table:

Title	Date	Agency	Type
Pope condemns excesses of genetic engineering	09/03/1998	AFP	text
Pope assails artificial reproduction techniques	03/04/2000	AFP	text
Pope slams cloning of human embryos	29/08/2000	AFP	text
Nobel winners unlocked progress towards treatment for Parkinson's	09/10/2000	AFP	text
Bush torn between science, religion over stem cell research	02/07/2001	AFP	text
Scientists affirm support for stem-cell research while debate rages	19/07/2001	AFP	text
Pope urges US to help reduce gap between rich and poor	23/07/2001	AFP	text
Pope warns Bush on death penalty, stem cell research	23/07/2001	AFP	text
Pope warns Bush on death penalty, stem cell research	23/07/2001	AFP	text
Vatican welcomes US vote to ban human cloning	01/08/2001	AFP	text
Bush to make decision on stem cells late Thursday	09/08/2001	AFP	text
Bush to announce decision on stem cells late Thursday	09/08/2001	AFP	text

Expressing information need



I' d like information about the relation of man-made things and GMO

From the perspective of a historian this might be expressed as:

The logo for Papyrus, featuring a stylized grid of colored squares above the word "papyrus" in a lowercase, sans-serif font.

papyrus

artifacts "genetically modified organism"

Selection of entities of interest

Concepts for query: artifacts "genetically modified organism"

Related Concepts	Selected				
<ul style="list-style-type: none">Historiographical issue<ul style="list-style-type: none">ArtifactsAward<ul style="list-style-type: none">Louisa Gross Horwitz prizeBiological object<ul style="list-style-type: none">Organism<ul style="list-style-type: none">Genetically modified organismCloned organism<ul style="list-style-type: none">Injaz<u>Dolly the Sheep</u>Tetra, the moriActivity<ul style="list-style-type: none">Creation<ul style="list-style-type: none">Dolly's creationActor / Agent<ul style="list-style-type: none">Person<ul style="list-style-type: none">Ian WilmutKeith H. S. CampDomain<ul style="list-style-type: none">Biotechnology domain	<p>Dolly the Sheep (remove)</p> <table border="1"><tbody><tr><td>has definition</td><td>"Dolly (5 July 1996 - 14 February 2003) was a female domestic sheep remarkable in being the first mammal to be cloned from an adult somatic cell, using the process of nuclear transfer. She was cloned by Ian Wilmut, Keith Campbell and colleagues at the Roslin Institute near Edinburgh in Scotland. She was born on 5 July 1996 and she lived until the age of six. She has been called "the world's most famous sheep" by sources including BBC News and Scientific American."</td></tr><tr><td>was brought into existence by</td><td>"Dolly's creation"</td></tr></tbody></table>	has definition	"Dolly (5 July 1996 - 14 February 2003) was a female domestic sheep remarkable in being the first mammal to be cloned from an adult somatic cell, using the process of nuclear transfer. She was cloned by Ian Wilmut, Keith Campbell and colleagues at the Roslin Institute near Edinburgh in Scotland. She was born on 5 July 1996 and she lived until the age of six. She has been called "the world's most famous sheep" by sources including BBC News and Scientific American."	was brought into existence by	"Dolly's creation"
has definition	"Dolly (5 July 1996 - 14 February 2003) was a female domestic sheep remarkable in being the first mammal to be cloned from an adult somatic cell, using the process of nuclear transfer. She was cloned by Ian Wilmut, Keith Campbell and colleagues at the Roslin Institute near Edinburgh in Scotland. She was born on 5 July 1996 and she lived until the age of six. She has been called "the world's most famous sheep" by sources including BBC News and Scientific American."				
was brought into existence by	"Dolly's creation"				

Search

Retrieval of related News Items

- ▶ themes
- ▶ concepts
- ▶ location
- ▶ organisation
- ▶ person
- ▶ event



World's first cloned dog is genuine, SKorea panel says

Date: Tue Jan 10 02:30:00 EET 2006 [View Metadata](#)

Language: english Main Theme: Cloning



Iran announces birth of first live cloned sheep

Date: Sat Sep 30 10:47:19 EEST 2006 [View Metadata](#)

Language: english Main Theme: Cloning



US biotech company offers to clone man's best friend

Date: Wed May 21 14:43:06 EEST 2008 [View Metadata](#)

Language: english Main Theme: Cloning



US biotech company offers to clone man's best friend

Date: Wed May 21 14:43:08 EEST 2008 [View Metadata](#)

Language: english Main Theme: Cloning



Scientists say they have created first cloned embryo from primate

Date: Wed Nov 14 11:39:07 EET 2007 [View Metadata](#)

Language: english Main Theme: Cloning

Conclusions

- ▶ **Papyrus current state**

- ▶ Ongoing project
- ▶ Feedback and contributions are welcome

Akrivi Katifori, vivi@di.uoa.gr

Aristotle Tympas, tympas@phs.uoa.gr

Jan Korsten, J.W.A.Korsten@tue.nl

- ▶ **Papyrus potential applications**

- ▶ Education of history
- ▶ Support for journalists
- ▶ Contextualization of the news content
- ▶ Dynamic system for exploring digital archives

Thank you for listening!