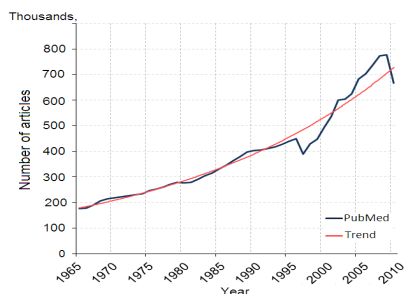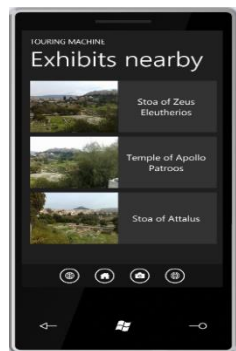# Human Language Technology for Cultural Heritage and Digital Humanities

Ion Androutsopoulos

Dept. of Informatics, Athens Univ. of Economics and Business

Digital Curation Unit, IMIS, Research Centre "Athena"

http://www.aueb.gr/users/ion/

Natural language generation and spoken dialogues in the INDIGO project

Project info:
www.ics.forth.gr/indigo
Video available at:
nlp.cs.aueb.gr/projects.html

Video recorded at the Foundation of the Hellenic World (FHW). Copyright of monument images: FHW. The descriptions of the monuments are generated from an OWL ontology based on information provided by FHW. INDIGO was an IST FP6 project funded by the European Union.

Authoring tool of IST M-PIRO. Developed at NCSR "Demokritos". Similar Protégé plug-in also available.
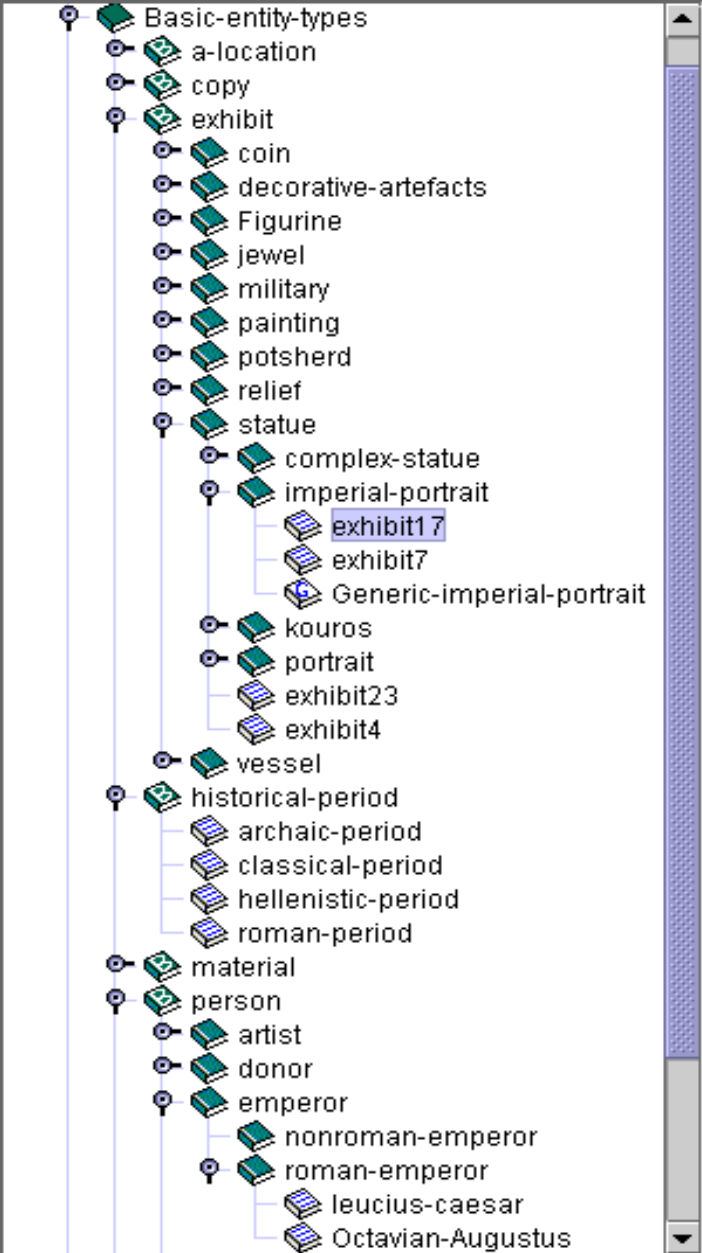
M-PIRO authoring tool --- C:\ion\projects\m-piro\software\java\authoring\authoring_v3_0_evaluation\mpiro_v2_...

Φάκελος   Επιλογές   Βοήθεια

ΤΥΠΟΙ ΧΡΗΣΤΩΝ   ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ   ΛΕΞΙΚΟ

NEO
ΑΝΟΙΓΜΑ
ΑΠΟΘ/ΣΗ
ΕΞΑΓΩΓΗ
ΕΞΟΔΟΣ

Υπόμνημα

Βασικοί
Τύποι

Τύποι
Οντοτήτων

Οντότητες

Γενικές
Οντότητες

Τύποι
Δεδομένων

- Basic-entity-types
  - a-location
  - copy
  - exhibit
    - coin
    - decorative-artefacts
    - Figurine
    - jewel
    - military
    - painting
    - potsherd
    - relief
    - statue
      - complex-statue
      - imperial-portrait
        - exhibit17
        - exhibit7
        - Generic-imperial-portrait
      - kouros
      - portrait
      - exhibit23
      - exhibit4
    - vessel
  - historical-period
    - archaic-period
    - classical-period
    - hellenistic-period
    - roman-period
  - material
  - person
    - artist
    - donor
    - emperor
      - nonroman-emperor
      - roman-emperor
        - leucius-caesar
        - Octavian-Augustus

Γλωσσικά-ανεξάρτητα πεδία του "exhibit17"

Πεδία	Τιμές
entity-id	exhibit17
type	imperial-portrait
images	<Augustus.jpg>
creation-period	roman-period
current-location	archaeological-thessaloniki
location-found	
original-location	Temple-Sarapis
made-of	marble
exhibit-height	Select a "Dimension"
copy-info	Select a "copy"
donate-by	Select a "donor"
sculpted-by	Select a "sculptor"
exhibit-shows	Select multiple .....
creation-time	between (17 year AD) (34 year AD)

Portait of Octavian Augustus

This exhibit is an imperial portrait; it was created during the roman period and it dates from between 17 and 34 A.D. It depicts Octavian Augustus and it is made from marble. This is one of the best examples of a propaganda statue.

_____

No notes

Natural language generation in virtual museums

Generating exhibit descriptions from OWL ontologies in Second Life (secondlife.com).
Video available at: nlp.cs.aueb.gr/projects.html

Information on your mobile phone

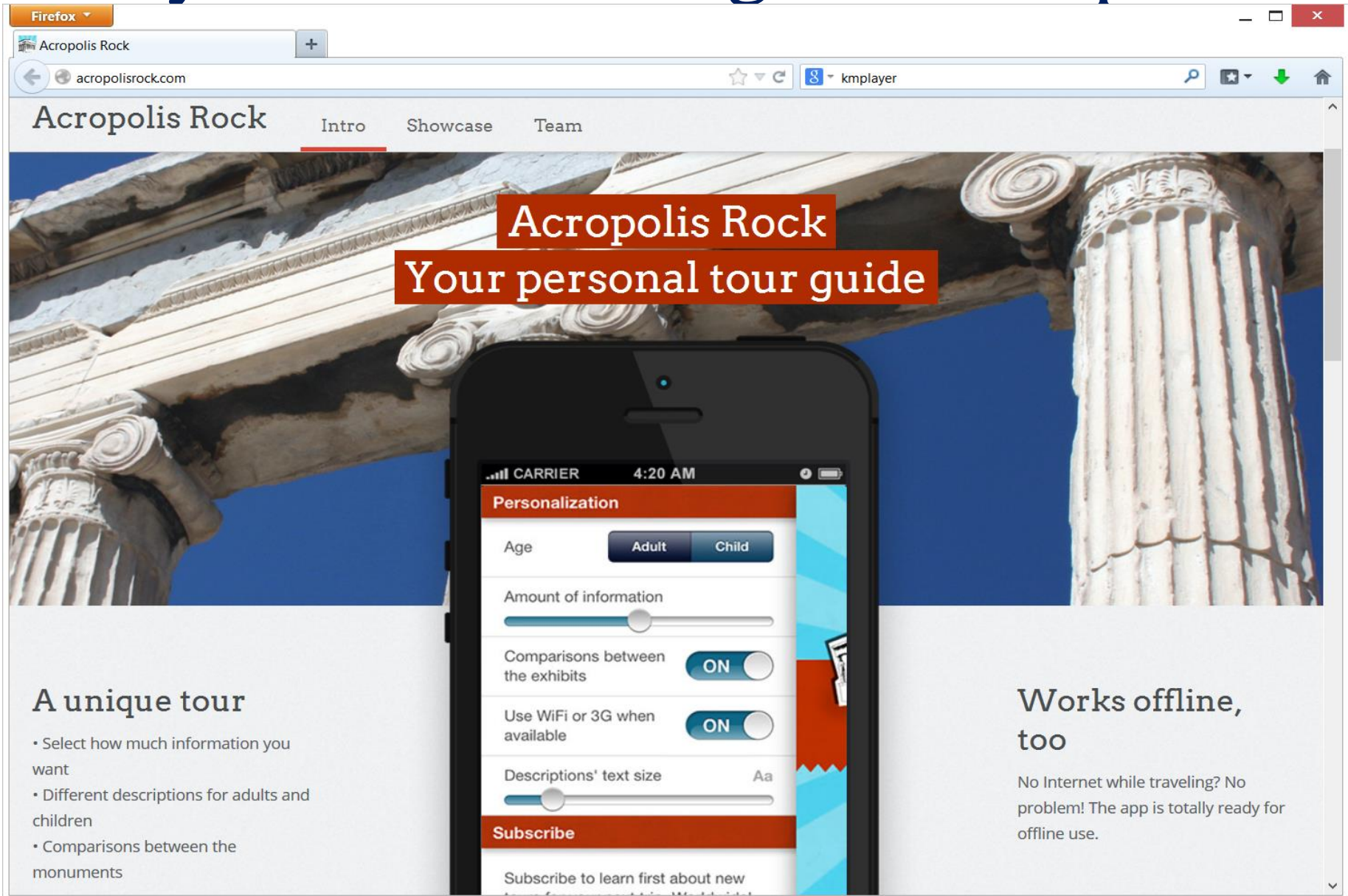Touring Machine (G. Karakatsiotis and V. Pterneas). Winners of the Greek Microsoft Imagine Cup 2010, winners of the International Imagine Cup 2011 (interoperability category). Texts generated from an OWL ontology of the Agora of Athens, created during INDIGO (www.ics.forth.gr/indigo) based on information of the Foundation for Hellenic World (www.fhw.gr/fhw).

Try it when visiting the Acropolis

Biomedical articles per year
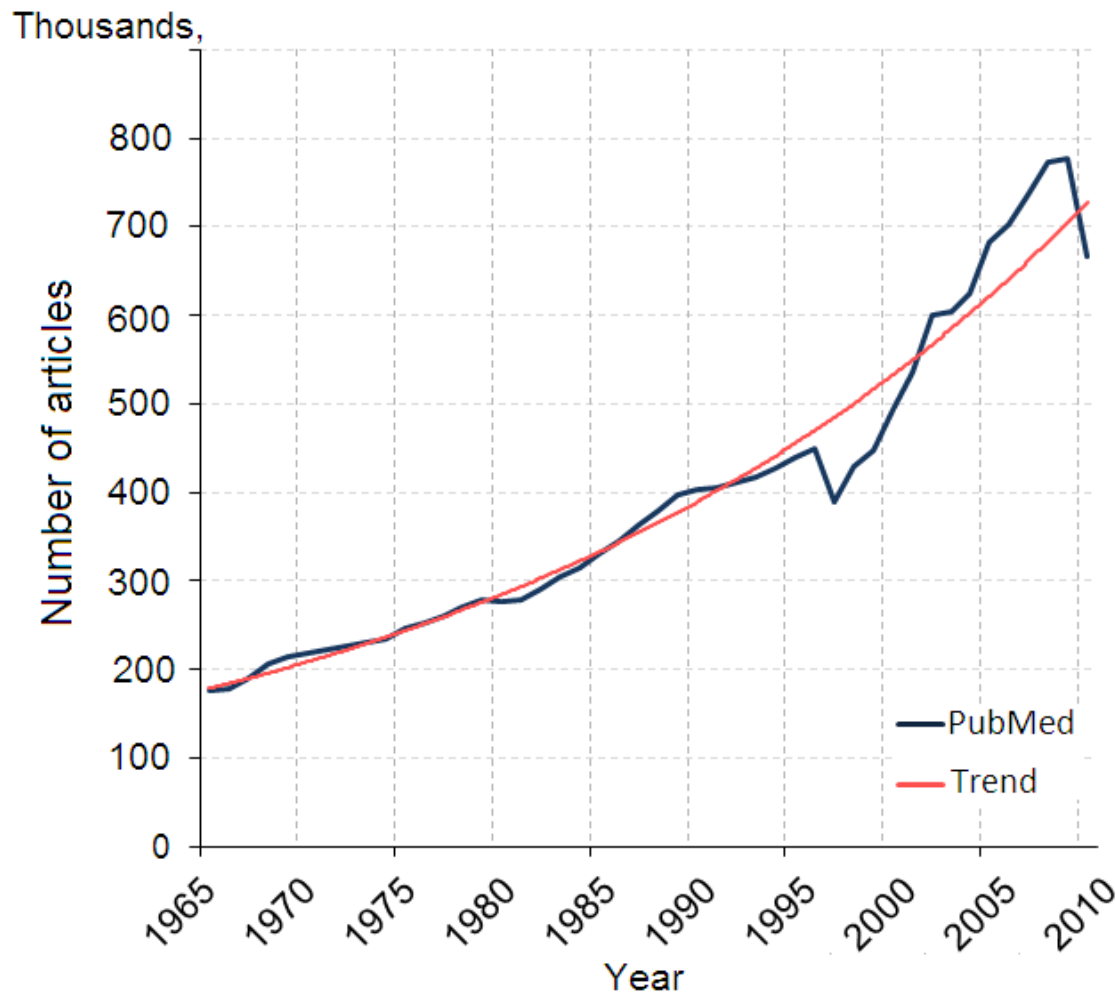
- In 2011: approx. 20 000 new PubMed articles *per week!*

Not only texts: ontologies, linked data, …
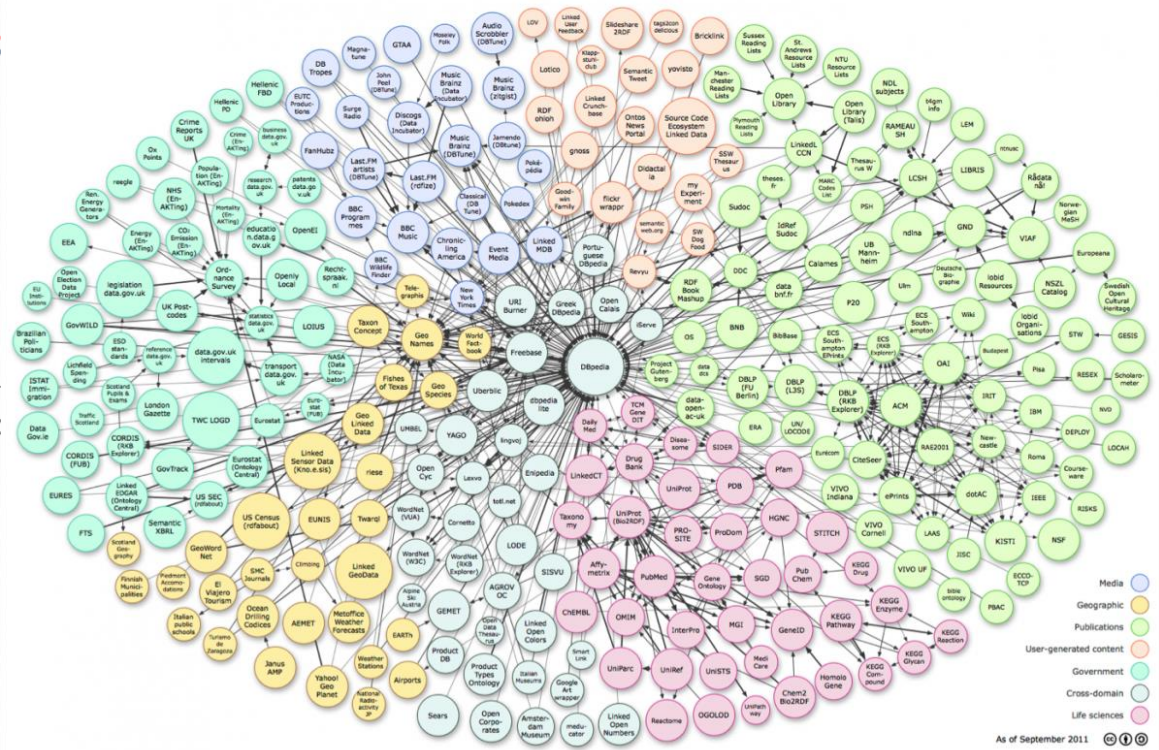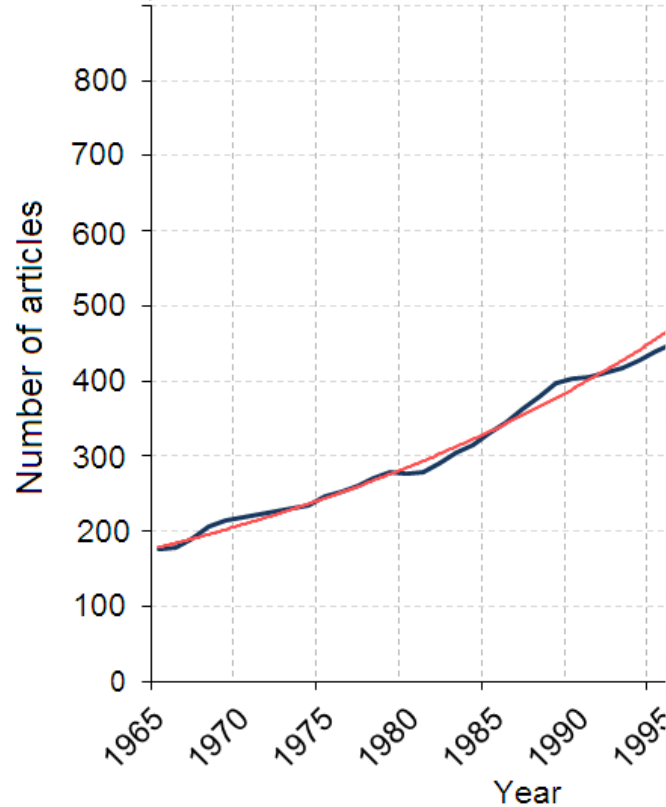
Questions of biomedical experts

- **Factoid question:** *"What is the methyl donor of DNA (cytosine-5)-methyltransferases?"*

  - **Exact answer**: *"S-adenosyl-L-methionine"*

  - **Ideal answer**: *"S-adenosyl-L-methionine (AdoMet, SAM) is the methyl donor of DNA (cytosine-5)-methyltransferases. DNA (cytosine-5)-methyltransferases catalyze the transfer of a methyl group from S-adenosyl-L-methionine to the C-5 position of cytosine residues in DNA."*

Questions of biomedical experts (II)

- **List question:** *"Which species may be used for the biotechnological production of itaconic acid?"*
  - **Exact answer**: { *"Aspergillus terreus", "Aspergillus niger", "Ustilago maydis"* }
  - **Ideal answer**: "*In 1955, the production of itaconic acid was firstly described for Ustilago maydis. Some Aspergillus species, like A. itaconicus and A. terreus, show the ability to synthesize this organic acid and A. terreus can secrete significant amounts to the media. Itaconic acid is mainly supplied by biotechnological processes with the fungus Aspergillus terreus. Cloning of the cadA gene into the citric acid producing fungus A. niger showed that it is possible to produce itaconic acid also in a different host organism.*"

What is the methyl donor of DNA (cytosi...

PubMed.gov
US National Library of Medicine
National Institutes of Health

[PubMed ▾] What is the methyl donor of DNA (cytosine-5)-methyltransferases? ⓧ **Search**

RSS   Save search   Advanced   Help

Show additional filters

Display Settings: ⌄ Summary, 20 per page, Sorted by Recently Added   Send to: ⌄   Filters: **Manage Filters**

⭐ Did you mean: _what is the methyl donor of dna (cytosine-5)_ **methyltransferases** (34 items)

New feature
Try the new Display Settings option - **Sort by Relevance**

Article types
Review
More ...

Results: 1 to 20 of 34   << First < Prev Page [1] of 2 Next > Last >>

Text availability
Abstract available
Free full text available
Full text available

☐ 1. **Cytosine-to-uracil deamination by SssI DNA methyltransferase.**
Stier I, Kiss A.
PLoS One. 2013 Oct 21;8(10):e79003. doi: 10.1371/journal.pone.0079003. eCollection 2013.
PMID: 24205358 [PubMed - in process]   **Free PMC Article**
Related citations

(methyl[All Fields] AND ("tissue donors"[MeSH Terms] OR ("tissue"[All Fields] AND "donors"[All Fields]) OR "tissue donors"[All Fields] OR

Search

See more...

Publication dates
5 years
10 years
Custom range...

☐ 2. **Long-term exposure to cigarette smoke extract induces hypomethylation at the RUNX3 and IGF2-H19 loci in immortalized human urothelial cells.**
Chen LM, Nergard JC, Ni L, Rosser CJ, Chai KX.
PLoS One. 2013 May 28;8(5):e65513. doi: 10.1371/journal.pone.0065513. Print 2013.
PMID: 23724145 [PubMed - indexed for MEDLINE]   **Free PMC Article**
Related citations

Species
Humans
Other Animals

☐ 3. **Diet influences expression of autoimmune-associated genes and disease severity by epigenetic mechanisms in a transgenic mouse model of lupus.**
Strickland FM, Hewagama A, Wu A, Sawalha AH, Delaney C, Hoeltzel MF, Yung R, Johnson K, Mickelson B, Richardson BC.
Arthritis Rheum. 2013 Jul;65(7):1872-81. doi: 10.1002/art.37967.
PMID: 23576011 [PubMed - indexed for MEDLINE]
Related citations

Clear all

Show additional filters

☐ 4. **The 'golden age' of DNA methylation in neurodegenerative diseases.**
Fuso A.
Clin Chem Lab Med. 2013 Mar 1;51(3):523-34. doi: 10.1515/cclm-2012-0618. Review.
PMID: 23183753 [PubMed - indexed for MEDLINE]
Related citations

☐ 5. **Wilson's disease: changes in methionine metabolism and inflammation affect global DNA methylation in early liver disease**

What is BioASQ?

A **systems competition** funded by the **EU (FP7)**.

Task A: Hierarchical text classification

- Organizers distribute **new unclassified PubMed articles**.
- Participating systems assign **MeSH terms** to the articles.
- **Evaluation** based on annotations of **PubMed curators**.

Task B: IR, QA, summarization, ...

- Organizers distribute **English biomedical questions**.
- Participating systems provide: relevant **articles**, **snippets**, **concepts**, **triples**, "**exact**", "**ideal**" **answers**.
- **Evaluation**: both **automatic** (GMAP, MRR, ROUGE etc.) and **manual** (by biomedical experts).

Similar Needs in Digital Humanities?

- Would it be useful to organize a **similar challenge** for **digital humanities**?

- **How** do **researchers** of **digital humanities search**?

- **Where** do they search? Which **repositories**?

- Collect **questions** reflecting real needs plus **gold answers** (articles, snippets, "exact", "ideal" answers)?

- For which **subfields**?

- **Funding**?

- **Contact me** if you are interested and/or know some of the answers...