



# BioASQ

A challenge on large-scale biomedical semantic indexing and question answering

[www.bioasq.org](http://www.bioasq.org)

John Pavlopoulos

NLP Group, Department of Informatics,  
Athens University of Economics and Business, Greece



Intelligent Information Management  
Targeted Competition Framework  
ICT-2011.4.4(d)



# 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).

# 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.”*

# More info about BioASQ

**Questions** (300+500) and **gold reference** answers prepared by **biomedical experts** from around Europe.

- ▶ Using tools/infrastructure developed by BioASQ.
- ▶ The questions reflect the **real needs** of the experts.

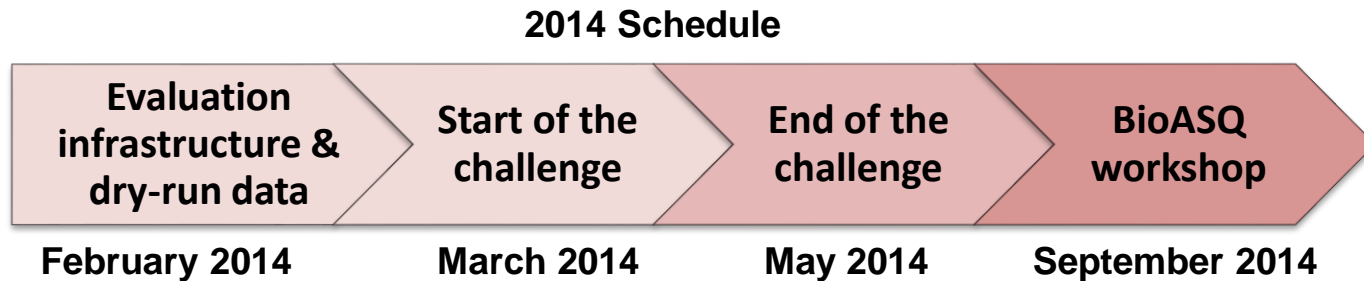
**Data sources** include both **text and structured** info.

- ▶ PubMed articles and MeSH.
- ▶ Gene Ontology, UniProt, Jochem, Disease Ontology.

BioASQ **datasets, infrastructure, evaluation services** etc. available **beyond the end** of the project.

- ▶ Plus **social net** to help extend data.

# Get involved!



- ▶ The official challenge is almost over, but...
- ▶ **Task A** will **continue to run** beyond the end of the project.
- ▶ An **oracle** for **Task B** will also remain **available**.
- ▶ Attend the **BioASQ workshop** at **CLEF '14** (15-18/9).
- ▶ You can **submit papers** to the workshop **even if you did not participate** in the official challenge.

# Get in touch!

**BioASQ workshop @CLEF (Sheffield, deadline June 15.)**

Visit **[www.bioasq.org](http://www.bioasq.org)**

Follow **@BioASQ**

# Project Consortium

1. National Centre for Scientific Research “Demokritos” - NSCR “D” (EL)
2. Transinsight GmbH – TI (D)
3. Université Joseph Fourier- UJF (F)
4. University Leipzig - ULEI (D)
5. Université Pierre et Marie Curie Paris 6 – UPMC (F)
6. Athens University of Economics and Business – Research Centre – AUEB-RC (EL)

