

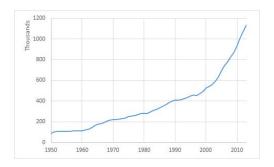
#### **BioASQ**

Project Vision and Overall approach

George Paliouras and Anastasia Krithara

paliourg@iit.demokritos.gr, akrithara@iit.demokritos.gr

### Vision



- 2 articles published in biomedical journals every minute!
- Make sure this knowledge is used to the benefit of patients
- Need to make it accessible to biomedical experts
- Search is not effective enough
- Push research in automated answering of questions
- ► A challenge for such systems can achieve a multiplying effect

### What is BioASQ

- BioASQ initiated a series of challenges on biomedical semantic indexing and question answering (QA).
- ► Participants are required to index semantically content from large-scale biomedical resources (e.g. MEDLINE) and/or
- to assemble data from multiple heterogeneous sources (e.g. scientific articles, knowledge bases, databases)
- to compose informative answers to biomedical natural language questions.

#### **BioASQ Tasks**

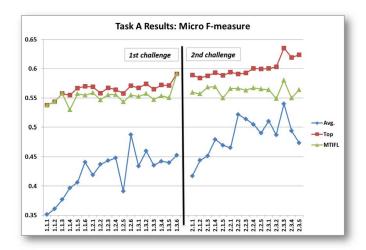
#### Task A: Hierarchical text classification

- Organizers distribute new unclassfied MEDLINE articles.
- ▶ Participants assign **MeSH terms** to the articles.
- Evaluation based on annotations of MEDLINE curators.

#### Task B: IR, QA, summarization

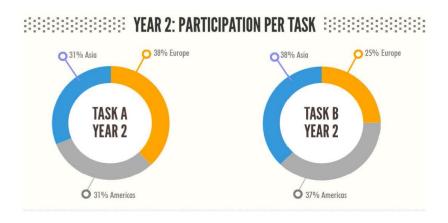
- Organizers distribute English biomedical questions.
- Participants provide: relevant articles, snippets, concepts, triples (phase A) and 'exact' answers, 'ideal' answers (phase B).
- Evaluation: both automated and manual.

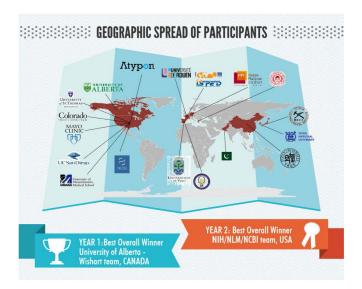
## Advancing the SoA in large-scale semantic indexing and QA.



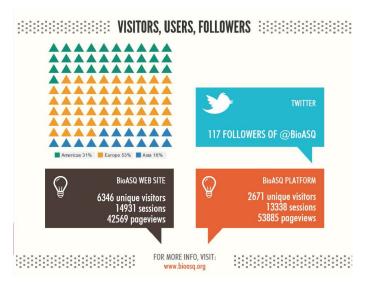














## BioASQ as a reference point in biomedical QA

- Attraction of key players in the field
  - ▶ Challenge Participants
  - Advisory board members
- BioASQ Social network
- Benchmark sets available
- NLM Statement



#### BioASQ Infrastructure

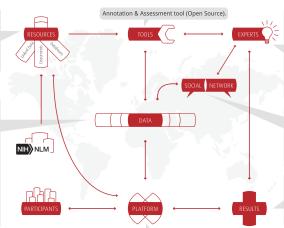
Data sources include both text and structured info

 MEDLINE articles.
 MeSH, Gene Ontology, UniProt, Jochem, Disease Ontology.

Task A: More than 10,000,000 articles with MeSH headings available for training. New test batches available every week.

Task B: More than 800 English questions, plus gold relevant documents, snippets, concepts, triples, "exact", and "ideal" answers.

Benchmark datasets/oracles available beyond the end of the project.



Online platform that enables participants to enter the BioASQ challenge. Infrastructure available beyond the end of the project.

Biomedical experts, responsible for the formulation of questions, the identification of gold relevant documents, snippets, concepts, triples, the composition of gold "exact" and "ideal" answers, and the manual evaluation of system responses.

A social network allows biomedical experts to follow and comment upon benchmark questions and gold answers formulated by their peers.

Automated & manual assessment of system responses. Oracles available for objective retrospective benchmarking beyond the end of the project.

## Future steps

- ▶ BioASQ in 2015
  - Third challenge is already running
  - Part of CLEF 2015



# Visit www.bioasq.org Platform participants-area.bioasq.org Follow @BioASQ

