# T4: Natural Language Processing in Watson

## ABSTRACT

Open domain Question Answering (QA) is a long standing research problem. Recently, IBM took on this challenge in the context of the Jeopardy! game. Jeopardy! is a well-known TV quiz show that has been airing on television in the United States for more than 25 years. It pits three human contestants against one another in a competition that requires answering rich natural language questions over a very broad domain of topics. The development of a system able to compete to grand champions in the Jeopardy! challenge led to the design of the DeepQA architecture and the implementation of Watson. The DeepQA project shapes a grand challenge in Computer Science that aims to illustrate how the wide and growing accessibility of natural language content and the integration and advancement of Natural Language Processing, Information Retrieval, Machine Learning, Knowledge Representation and Reasoning, and massively parallel computation can drive open-domain automatic Question Answering technology to a point where it clearly and consistently rivals the best human performance. Natural Language Processing (NLP) plays a crucial role in the overall Deep QA architecture. It allows to "make sense" of both question and unstructured knowledge contained in the large corpora where most of the answers are located. That's why we decided to focus this tutorial on the NLP technology adopted by Watson and on how it fits in the general Deep QA architecture.

## OUTLINE

The course is structured in two modules (1.5h each), described below.

1. **Module 1 An overview of the Deep QA project**

   This lesson provides a general introduction to the Deep QA project, addressing the following topics

   1. Analysis of the Jeopardy Task
   2. The Deep QA architecture
   3. Development Cycle

2. **Module 2 Natural Language Processing in Watson**

   This lesson focuses on the NLP technology implemented in Watson, highlighting the advancements with respect to state of the art techniques in Question Answering. In particular:

1. Role of NLP in the Deep QA architecture
2. UIMA
3. Passage Scoring
4. Prismatic

## BIOS

Instructors will be Research Staff Members from the Algorithms team developing Watson at IBM Research. Depending on availability for the specific day of the tutorial, they will be one or more from the following list:

**Alfio M. Gliozzo**
19 Skyline Drive, Hawthorne, NY 10532
phone: +1-917-744-4351
gliozzo--AT--us.ibm.com
https://researcher.ibm.com/researcher/view.php?person=us-gliozzo

Dr. Alfio Gliozzo is Research Staff Member at IBM Research, where he is part of the Deep QA team. His main research focus is Domain Adaptation of Question Answering systems using machine learning and knowledge acquisition. He is author of about 50 publications in the areas of Computational Linguistics, Information Retrieval, and Semantic Web.

**Aditya Kalyanpur**
19 Skyline Drive, Hawthorne, NY 10532
adityakal--AT--us.ibm.com
http://researcher.ibm.com/researcher/view.php?person=us-adityakal

Aditya Kalyanpur is a Research Staff Member at IBM Research. His research interests include knowledge representation & reasoning, ontologies, natural language processing, machine learning and statistical data mining. Aditya has been involved in the development of several core algorithms in Watson related to question analysis, evidence gathering and scoring, knowledge based inference, and answer merging and ranking.

**James Fan**
19 Skyline Drive, Hawthorne, NY 10532
fanj--AT--us.ibm.com
https://researcher.ibm.com/researcher/view.php?person=us-fanj

James Fan is a research staff member at IBM Research. James has been working in the area of question answering since 2003, and has published on subjects in knowledge representation, reasoning, natural language processing and machine learning. James is currently working the DeepQA project.