“Housekeeping”

Twitter: #ACMLearning

- Welcome to today’s ACM SIGAI Learning Webinar sponsored by the ACM Special Interest Group on Artificial Intelligence, “On the Evolution of NLP, QA, and IE, and Current Research and Commercial Trends.” The presentation starts at the top of the hour and lasts 60 minutes. Audio and video will automatically play throughout the event. On the bottom panel you’ll find a number of widgets.

- If you are experiencing any problems/issues, **refresh** your console by pressing the **F5** key on your keyboard in **Windows**, **Command + R** if on a **Mac**, or refresh your browser if you’re on a mobile device; or close and re-launch the presentation. You can also view the Webcast Help Guide, by clicking on the “Help” widget in the bottom dock.

- To control volume, adjust the master volume on your computer. If the volume is still too low, use headphones.

- **If you think of a question during the presentation, please type it into the Q&A box and click on the submit button. You can start submitting your questions into the queue now and continue to enter more questions, as you come up with them, throughout the presentation.**

- This session is being recorded and will be archived for on-demand viewing in the next 1-2 days. You will receive an automatic email notification when it is available.

ACM and SIGAI Highlights

- Learning Center tools for professional development: http://learning.acm.org
- Popular publications:
  - Flagship *Communications of the ACM (CACM)* magazine: http://cacm.acm.org/
  - *ACM Queue* magazine for practitioners: http://queue.acm.org/
- ACM Digital Library, the world's most comprehensive database of computing literature: http://dl.acm.org.
- International conferences that draw leading experts on a broad spectrum of computing topics: http://www.acm.org/conferences.
- Prestigious awards, including the ACM A.M. Turing and ACM Prize in Computing: http://awards.acm.org
- And much more… http://www.acm.org.

- The **ACM Special Interest Group on Artificial Intelligence (SIGAI)** is made up of academic and industrial researchers, practitioners, software developers, end users, and students.
- Our goal is to:
  - Promote and support the growth and application of AI principles and techniques throughout computing
  - Sponsor or co-sponsor high-quality, AI-related conferences
  - Publish the quarterly newsletter AI Matters and its namesake blog
  - Organize the Career Network and Conference (SIGAI CNC) for early-stage researchers in AI
  - Sponsor recognized AI awards
  - Support important journals in the field
  - Provide scholarships to student members to attend conferences
  - Promote AI education and publications through various forums and the ACM digital library
- AI is increasingly an interdisciplinary area, and your membership supports these goals. You can join us today by going to the link on your screen or googling SIGAI. Both members and non-members are welcome to join our mailing list to receive timely announcements of interest to researchers and practitioners.

“Housekeeping”

Twitter: #ACMLearning

- Welcome to today’s ACM SIGAI Learning Webinar sponsored by the ACM Special Interest Group on Artificial Intelligence, “On the Evolution of NLP, QA, and IE, and Current Research and Commercial Trends.” The presentation starts at the top of the hour and lasts 60 minutes. Audio and video will automatically play throughout the event. On the bottom panel you’ll find a number of widgets.

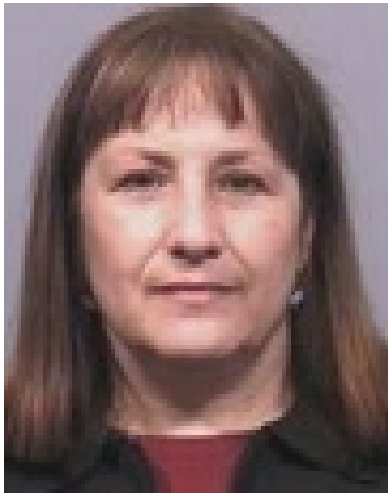
- If you are experiencing any problems/issues, **refresh** your console by pressing the **F5** key on your keyboard in **Windows**, **Command + R** if on a **Mac**, or refresh your browser if you’re on a mobile device; or close and re-launch the presentation. You can also view the Webcast Help Guide, by clicking on the “Help” widget in the bottom dock.

- To control volume, adjust the master volume on your computer. If the volume is still too low, use headphones.

- **If you think of a question during the presentation, please type it into the Q&A box and click on the submit button. You can start submitting your questions into the queue now and continue to enter more questions, as you come up with them, throughout the presentation.**

- This session is being recorded and will be archived for on-demand viewing in the next 1-2 days. You will receive an automatic email notification when it is available.

Welcome



Left to right:
Dan Moldovan, **Speaker:** Professor at the University of Texas at Dallas and Founder of Lymba Corporation
Plamen Petrov, Director of Cognitive Technology, KPMG LLP;
Rosemary Paradis, Principal Research Engineer at Leidos Health and Life Sciences.

*If you think of a question during the presentation, please type it into the Q&A box and click on the submit button. Since this presentation is fully question-based, you can start submitting your questions into the queue now and continue to enter more questions throughout the presentation.

On the evolution of NLP, QA and IE, and current research and commercial trends

Dan Moldovan

University of Texas
at Dallas

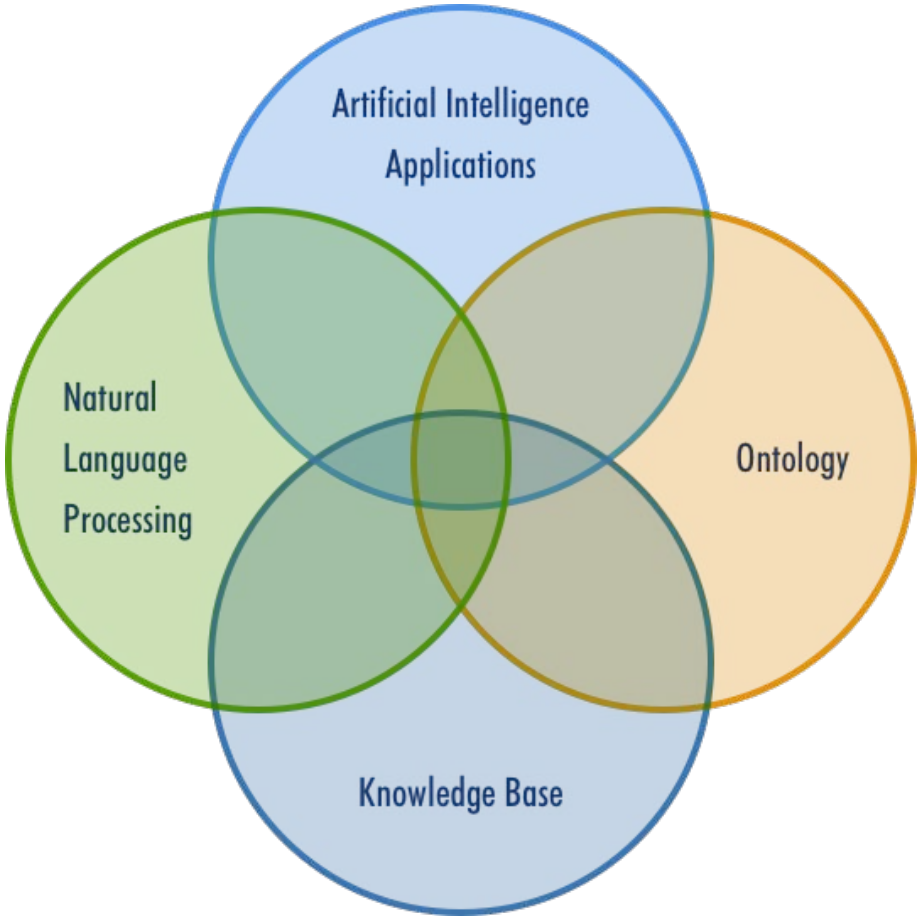
Outline

- Some definitions
- Areas of NLP
- Generic NLP pipeline
- Brief History of NLP, QA and IE
- Current research trends
- Current commercial trends

Definitions - NLP

- **Natural Language Processing** is a technology that creates and implements computer models for the purpose of performing various natural language tasks. It is used for building NL interfaces to databases, machine translation, and others.
- NLP was and remains one of the most important fields of **artificial intelligence**
- NLP-based AI applications

NLP-based AI applications



Definitions - QA

- **Question Answering** is the task of automatically responding to natural-language questions with relevant answers drawn from a database, knowledge base, or unstructured collection of natural language documents
- QA has been most useful so far for answering factoid-type questions
- QA is increasingly being used to build advanced chatbots

Definitions - IE

- **Information Extraction** is the task of extracting concepts, events and relations from text, thus transforming unstructured data into structured information.
- IE helps us answer questions like:

  Who did what to whom, when, where, why ?

Related areas

- NLP is a difficult, and largely unsolved problem. One reason for this is its **multidisciplinary** nature:
  - **Linguistics** : How words, phrases, and sentences are formed.
  - **Psycholinguistics** : How people understand and communicate using human language.
  - **Computational linguistics:** Deals with models and computational aspects of NL (e.g. algorithms).

Related areas

- **Philosophy**: relates to the semantics of language; notion of meaning, how words identify objects. NLP requires considerable knowledge about the world.
- **Computer science**: model formulation and implementation using modern methods.
- **Artificial intelligence**: issues related to knowledge representation and reasoning.
- **Statistics:** many NLP problems are modeled using probabilistic models.
- **Machine learning:** automatic learning of rules and procedures based on lexical, syntactic and semantic features.
- **NL Engineering**: implementation of large, realistic systems. Modern software development methods play an important role.

A Typical NLP Pipeline

Applications of NLP

- Sentiment Analysis
- Text Summarization
- Textual Entailment
- Information Extraction
- Topic Segmentation
- Question Answering
- Semantic comparison of two documents
- Ontology Building
- Event Detection and Reasoning
- Regulatory Compliance

Applications of NLP

- Trend Analysis and Prediction
- Risk Management; Risk Assessment
- Decision Making and Evidence Support
- Customer Profiles and Business Intelligence
- CRM, call centers using text messages
- Translation
- Argumentation Mining

History of NLP: 1950's – 1960's

- 1950 – Alan Turing proposes the Turing test
- 1954 – First attempts at machine translation: Georgetown-IBM experiment
- 1956 – Dartmouth Workshop: "artificial intelligence" as a field of study, with NLP as one of its chief areas
- 1956 – Hierarchy of formal grammars (Chomsky)
- 1964 – First NLP systems: ELIZA (Weizenbaum), STUDENT (Bobrow)
- 1968 – SHRDLU (Winograd) virtual world NLUI

History of NLP: 1970's – 1980's

- 1972 – PARRY (Colby) chatterbot with personality
- 1975 – MARGIE (Schank) semantic parser for paraphrasing/translation
- 1978 – SAM (Cullingford) story understanding program
- 1980's – Machine learning enters NLP: decision trees, hidden Markov models for POS tagging, noisy channel theory
- 1980's – MUC (IE) begins: naval operations messages
- 1980's – Fifth Generation Computer Project in Japan – overpromised and under-delivered. Triggered an AI winter.

History of NLP: 1990's to today

- 1990's – Statistical NLP, parsed corpora: Penn Treebank
- 1990's – MUC (IE) continues: terrorism in Latin America, joint ventures, management changes, satellite launches
- 2000's – Semantic Web: RDF, SPARQL, OWL
- 2000's – AQUAINT I,II, III projects funded by iARPA
- 2000's – NIMD (iARPA), GALE (DARPA), KDD (iARPA)
- 2010's – IBM Watson wins Jeopardy
- 2010's – Big Data, word embeddings, deep learning
- 2010's – Attempts to understand implicatures
- 2010's – NLP applied to social media
- today – Unprecedented interest in NLP, AI across domains

Research Trends - NLU

- Document processing: document segmentation
- Semantics
  - Semantic parsers
  - Semantic calculus
- Discourse
  - Documents
  - Conversations (chatbots & dialog systems)

Research Trends - NLG

- As NLU problems are addressed, increasing attention is paid to NLG problems
- NLG applications
  - Summarization of large texts or document collections
  - Article/report generation from text, images, or other data
  - Question generation from textbooks
  - Answer generation for QA systems
  - Response generation for chatbots and computer game AI characters
  - Generation of creative works (poetry, prose fiction)

Research Trends - Techniques

- Increasing use of structured formats for semantics, discourse, pragmatics, sentiment, etc.
- Increased use of deep learning
  - Word embeddings (word2vec, GloVe)
  - Convolutional neural networks (CNNs) for extracting salient features
  - Recurrent neural networks (RNNs, LSTMs, GRUs) for modeling sequences of text
  - Generative adversarial networks (GANs) for generating human-like natural language output

Commercial Trends

- Systems for processing Big Data
- IE: converting unstructured text to actionable knowledge (structured data: RDF, XML, JSON)
- Creating/refining domain ontologies, Increased use of ontologies
- Enterprise Knowledge Graphs, RDF stores, graph databases
- Conversational agents: chatbots and dialog systems
- Human to Machine communication in addition to M2M
- QA: Information retrieval systems with natural language query

Commercial Trends

- Tools for rapid domain customization
- Decision and predictions as a service
- Document understanding combining text, tables, figures, images, drawings and graphs.
- Applications get smarter; NLP – enabled AI in areas such as contract processing, decision making, argumentation reasoning, etc.

Major Hurdles to Overcome

- NLP systems do not scale up well across domains. There is need for substantial amount of domain knowledge.
- Lack of understanding or agreement on a library of basic core capabilities; ie semantic comparison of documents, etc. on which applications can map into.
- The jump from sentence level NLP to document level is an enormous step and it takes time.
- NLP in native languages, lack of resources.
- Advanced NLP requires advanced Tools, and many of these remain proprietary and expensive.

ACM: The Learning Continues…

- Questions about this webcast? learning@acm.org
- ACM Learning Webinars (on-demand archive): http://webinar.acm.org
- SIGAI Website: https://sigai.acm.org
- SIGAI Newsletter – AI MATTERS: https://sigai.acm.org/aimatters/
- SIGAI Organization: https://sigai.acm.org/organization/advisory_board.html