Systems & Applications: Introduction

Ling 573
NLP Systems and Applications
April 1, 2014

Roadmap

- Motivation

- 573 Structure

- Question-Answering

- Shared Tasks

Motivation

- Information retrieval is very powerful
  - Search engines index and search enormous doc sets
  - Retrieve billions of documents in tenths of seconds

Motivation

- Information retrieval is very powerful
  - Search engines index and search enormous doc sets
  - Retrieve billions of documents in tenths of seconds

- But still limited!

Motivation

- Information retrieval is very powerful
  - Search engines index and search enormous doc sets
  - Retrieve billions of documents in tenths of seconds

- But still limited!
  - Technically – keyword search (mostly)

Motivation

- Information retrieval is very powerful
  - Search engines index and search enormous doc sets
  - Retrieve billions of documents in tenths of seconds

- But still limited!
  - Technically – keyword search (mostly)
  - Conceptually
    - User seeks information
      - Sometimes a web site or document

Motivation

- Information retrieval is very powerful
  - Search engines index and search enormous doc sets
  - Retrieve billions of documents in tenths of seconds

- But still limited!
  - Technically – keyword search (mostly)
  - Conceptually
    - User seeks information
      - Sometimes a web site or document
      - Very often, the answer to a question

Why Question-Answering?

- People ask questions on the web

Why Question-Answering?

- People ask questions on the web
  - Web logs:
    - Which English translation of the bible is used in official Catholic liturgies?
    - Who invented surf music?
    - What are the seven wonders of the world?

Why Question-Answering?

- People ask questions on the web
  - Web logs:
    - Which English translation of the bible is used in official Catholic liturgies?
    - Who invented surf music?
    - What are the seven wonders of the world?
  - 12-15% of queries

Why Question-Answering?

- People ask questions on the web
  - Web logs:
    - Which English translation of the bible is used in official Catholic liturgies?
    - Who invented surf music?
    - What are the seven wonders of the world?
  - 12-15% of queries

- Search sites (e.g., Google) beginning to include
  - Canonical factoids, esp. Wikipedia infobox data
    - Dates, conversions, birthdates

Why Question Answering?

- Answer sites proliferate:

Why Question Answering?

- Answer sites proliferate:
  - Top hit for 'questions' :

Why Question Answering?

- Answer sites proliferate:
  - Top hit for 'questions' : Ask.com

Why Question Answering?

- Answer sites proliferate:
  - Top hit for 'questions' : Ask.com
    - Also: Yahoo! Answers, wiki answers, Facebook,…
  - Collect and distribute human answers

Why Question Answering?

- Answer sites proliferate:
  - Top hit for 'questions' : Ask.com
    - Also: Yahoo! Answers, wiki answers, Facebook,...
  - Collect and distribute human answers
  - **Do I Need a Visa to Go to Japan?**

Why Question Answering?

- Answer sites proliferate:
  - Top hit for 'questions' : Ask.com
    - Also: Yahoo! Answers, wiki answers, Facebook,...
  - Collect and distribute human answers
  - **Do I Need a Visa to Go to Japan?**
    - **eHow.com**
    - Rules regarding travel between the United States and Japan are governed by both countries. Entry requirements for Japan are contingent on the purpose and length of a traveler's visit.
    - **Passport Requirements**
      - Japan requires all U.S. citizens provide a valid passport and a return on "onward" ticket for entry into the country. Additionally, the United States requires a passport for all citizens wishing to enter or re-enter the country.

Search Engines & QA

- *Who was the prime minister of Australia during the Great Depression?*

Search Engines & QA

- *Who was the prime minister of Australia during the Great Depression?*
  - Rank 1 snippet:
    - The conservative *Prime Minister of Australia*, Stanley Bruce

Search Engines & QA

- *Who was the prime minister of Australia during the Great Depression?*
  - Rank 1 snippet:
    - The conservative *Prime Minister of Australia*, Stanley Bruce

  - Wrong!
    - Voted out just before the Depression

Perspectives on QA

- TREC QA track (1999…)
  - Initially pure factoid questions, with fixed length answers
    - Based on large collection of fixed documents (news)
    - Increasing complexity: definitions, biographical info, etc
      - Single response

Perspectives on QA

- TREC QA track (~1999···)
  - Initially pure factoid questions, with fixed length answers
    - Based on large collection of fixed documents (news)
    - Increasing complexity: definitions, biographical info, etc
      - Single response

- Reading comprehension (Hirschman et al, 2000···)
  - Think SAT/GRE
    - Short text or article (usually middle school level)
    - Answer questions based on text
  - Also, 'machine reading'

Perspectives on QA

- TREC QA track (~1999···)
  - Initially pure factoid questions, with fixed length answers
    - Based on large collection of fixed documents (news)
    - Increasing complexity: definitions, biographical info, etc
      - Single response

- Reading comprehension (Hirschman et al, 2000···)
  - Think SAT/GRE
    - Short text or article (usually middle school level)
    - Answer questions based on text
  - Also, 'machine reading'

- And, of course, Jeopardy! and Watson

Natural Language Processing and QA

- Rich testbed for NLP techniques:

Natural Language Processing and QA

- Rich testbed for NLP techniques:
  - Information retrieval

Natural Language Processing and QA

- Rich testbed for NLP techniques:
  - Information retrieval
  - Named Entity Recognition

Natural Language Processing and QA

- Rich testbed for NLP techniques:
  - Information retrieval
  - Named Entity Recognition
  - Tagging

Natural Language Processing and QA

- Rich testbed for NLP techniques:
  - Information retrieval
  - Named Entity Recognition
  - Tagging
  - Information extraction

Natural Language Processing and QA

- Rich testbed for NLP techniques:
  - Information retrieval
  - Named Entity Recognition
  - Tagging
  - Information extraction
  - Word sense disambiguation

Natural Language Processing and QA

- Rich testbed for NLP techniques:
  - Information retrieval
  - Named Entity Recognition
  - Tagging
  - Information extraction
  - Word sense disambiguation
  - Parsing

Natural Language Processing and QA

- Rich testbed for NLP techniques:
  - Information retrieval
  - Named Entity Recognition
  - Tagging
  - Information extraction
  - Word sense disambiguation
  - Parsing
  - Semantics, etc..

Natural Language Processing and QA

- Rich testbed for NLP techniques:
  - Information retrieval
  - Named Entity Recognition
  - Tagging
  - Information extraction
  - Word sense disambiguation
  - Parsing
  - Semantics, etc..
  - Co-reference

Natural Language Processing and QA

- Rich testbed for NLP techniques:
    - Information retrieval
    - Named Entity Recognition
    - Tagging
    - Information extraction
    - Word sense disambiguation
    - Parsing
    - Semantics, etc..
    - Co-reference

- Deep/shallow techniques; machine learning

573 Structure

- Implementation:

573 Structure

- Implementation:
  - Create a factoid QA system

573 Structure

- Implementation:
    - Create a factoid QA system
        - Extend existing software components
        - Develop, evaluate on standard data set

573 Structure

- Implementation:
  - Create a factoid QA system
    - Extend existing software components
    - Develop, evaluate on standard data set

- Presentation:

573 Structure

- Implementation:
  - Create a factoid QA system
    - Extend existing software components
    - Develop, evaluate on standard data set

- Presentation:
  - Write a technical report
  - Present plan, system, results in class

573 Structure

- Implementation:
  - Create a factoid QA system
    - Extend existing software components
    - Develop, evaluate on standard data set

- Presentation:
  - Write a technical report
  - Present plan, system, results in class
  - Give/receive feedback

Implementation: Deliverables

- Complex system:
  - Break into (relatively) manageable components
  - Incremental progress, deadlines

Implementation: Deliverables

- Complex system:
  - Break into (relatively) manageable components
  - Incremental progress, deadlines

- Key components:
  - D1: Setup

Implementation: Deliverables

- Complex system:
  - Break into (relatively) manageable components
  - Incremental progress, deadlines

- Key components:
  - D1: Setup
  - D2: Baseline system, Passage retrieval

Implementation: Deliverables

- Complex system:
  - Break into (relatively) manageable components
  - Incremental progress, deadlines

- Key components:
  - D1: Setup
  - D2: Baseline system, Passage retrieval
  - D3: Question processing, classification

Implementation: Deliverables

- Complex system:
  - Break into (relatively) manageable components
  - Incremental progress, deadlines

- Key components:
  - D1: Setup
  - D2: Baseline system, Passage retrieval
  - D3: Question processing, classification
  - D4: Answer processing, final results

Implementation: Deliverables

- Complex system:
  - Break into (relatively) manageable components
  - Incremental progress, deadlines

- Key components:
  - D1: Setup
  - D2: Baseline system, Passage retrieval
  - D3: Question processing, classification
  - D4: Answer processing, final results

- Deadlines:
  - Little slack in schedule; please keep to time
  - Timing: ~12 hours week; sometimes higher

Presentation

- Technical report:
  - Follow organization for scientific paper
    - Formatting and Content

Presentation

- Technical report:
  - Follow organization for scientific paper
    - Formatting and Content

- Presentations:
  - 10-15 minute oral presentation for deliverables

Presentation

- Technical report:
  - Follow organization for scientific paper
    - Formatting and Content

- Presentations:
  - 10-15 minute oral presentation for deliverables
  - Explain goals, methodology, success, issues

Presentation

- Technical report:
  - Follow organization for scientific paper
    - Formatting and Content

- Presentations:
  - 10-15 minute oral presentation for deliverables
  - Explain goals, methodology, success, issues
  - Critique each others' work

Presentation

- Technical report:
  - Follow organization for scientific paper
    - Formatting and Content

- Presentations:
  - 10-15 minute oral presentation for deliverables
  - Explain goals, methodology, success, issues
  - Critique each others' work
  - Attend **ALL** presentations

Working in Teams

- Why teams?

Working in Teams

- Why teams?
  - Too much work for a single person
  - Representative of professional environment

Working in Teams

- Why teams?
  - Too much work for a single person
  - Representative of professional environment

- Team organization:
  - Form groups of 3 (possibly 2) people

Working in Teams

- Why teams?
  - Too much work for a single person
  - Representative of professional environment

- Team organization:
  - Form groups of 3 (possibly 2) people
  - Arrange coordination
  - Distribute work equitably

Working in Teams

- Why teams?
  - Too much work for a single person
  - Representative of professional environment

- Team organization:
  - Form groups of 3 (possibly 2) people
  - Arrange coordination
  - Distribute work equitably
    - All team members receive the same grade
      - End-of-course evaluation

First Task

- Form teams:
  - Email Glenn gslaydeni@uw.edu with the team list

Resources

- Readings:
  - Current research papers in question-answering

Resources

- Readings:
  - Current research papers in question-answering
  - Jurafsky & Martin/Manning & Schutze text
    - Background, reference, refresher

Resources

- Readings:
  - Current research papers in question-answering
  - Jurafsky & Martin/Manning & Schutze text
    - Background, reference, refresher

- Software:

Resources

- Readings:
  - Current research papers in question-answering
  - Jurafsky & Martin/Manning & Schutze text
    - Background, reference, refresher

- Software:
  - Build on existing system components, toolkits
    - NLP, machine learning, etc
    - Corpora, etc

Resources: Patas

- System should run on patas
  - Existing infrastructure
    - Software systems

    - Corpora

    - Repositories

Shared Task Evaluations

- Goals:
  - Lofty:

Shared Task Evaluations

- Goals:
  - Lofty:
    - Focus research community on key challenges
      - 'Grand challenges'

Shared Task Evaluations

- Goals:
  - Lofty:
    - Focus research community on key challenges
      - 'Grand challenges'

    - Support the creation of large-scale community resources
      - Corpora: News, Recordings, Video
      - Annotation: Expert questions, labeled answers,..

Shared Task Evaluations

- Goals:
  - Lofty:
    - Focus research community on key challenges
      - 'Grand challenges'

    - Support the creation of large-scale community resources
      - Corpora: News, Recordings, Video
      - Annotation: Expert questions, labeled answers,..

    - Develop methodologies to evaluate state-of-the-art
      - Retrieval, Machine Translation, etc

Shared Task Evaluations

- Goals:
  - Lofty:
    - Focus research community on key challenges
      - 'Grand challenges'

    - Support the creation of large-scale community resources
      - Corpora: News, Recordings, Video
      - Annotation: Expert questions, labeled answers,..

    - Develop methodologies to evaluate state-of-the-art
      - Retrieval, Machine Translation, etc

    - Facilitate technology/knowledge transfer b/t industry/acad.

Shared Task Evaluation

- Goals:
  - Pragmatic:

Shared Task Evaluation

- Goals:
  - Pragmatic:
    - Head-to-head comparison of systems/techniques
      - Same data, same task, same conditions, same timing

Shared Task Evaluation

- Goals:
  - Pragmatic:
    - Head-to-head comparison of systems/techniques
      - Same data, same task, same conditions, same timing
    - Centralizes funding, effort

Shared Task Evaluation

- Goals:
  - Pragmatic:
    - Head-to-head comparison of systems/techniques
      - Same data, same task, same conditions, same timing
    - Centralizes funding, effort
    - Requires disclosure of techniques in exchange for data

  - Base:

Shared Task Evaluation

- Goals:
  - Pragmatic:
    - Head-to-head comparison of systems/techniques
      - Same data, same task, same conditions, same timing
    - Centralizes funding, effort
    - Requires disclosure of techniques in exchange for data

  - Base:
    - Bragging rights

Shared Task Evaluation

- Goals:
  - Pragmatic:
    - Head-to-head comparison of systems/techniques
      - Same data, same task, same conditions, same timing
    - Centralizes funding, effort
    - Requires disclosure of techniques in exchange for data

  - Base:
    - Bragging rights
    - Government research funding decisions

Shared Tasks: Perspective

- Late '80s-90s:

Shared Tasks: Perspective

- Late '80s-90s:
  - ATIS: spoken dialog systems
  - MUC: Message Understanding: information extraction

Shared Tasks: Perspective

- Late '80s-90s:
  - ATIS: spoken dialog systems
  - MUC: Message Understanding: information extraction

- TREC (Text Retrieval Conference)
  - Arguably largest (often >100 participating teams)
  - Longest running (1992-current)
  - Information retrieval (and related technologies)
    - Actually hasn't had 'ad-hoc' since ~2000, though
  - Organized by NIST

TREC Tracks

- Track: Basic task organization

TREC Tracks

- Track: Basic task organization

- Previous tracks:
  - Ad-hoc – Basic retrieval from fixed document set

TREC Tracks

- Track: Basic task organization

- Previous tracks:
  - Ad-hoc – Basic retrieval from fixed document set
  - Cross-language – Query in one language, docs in other
    - English, French, Spanish, Italian, German, Chinese, Arabic

TREC Tracks

- Track: Basic task organization

- Previous tracks:
  - Ad-hoc – Basic retrieval from fixed document set
  - Cross-language – Query in one language, docs in other
    - English, French, Spanish, Italian, German, Chinese, Arabic
  - Genomics

TREC Tracks

- Track: Basic task organization

- Previous tracks:
  - Ad-hoc – Basic retrieval from fixed document set
  - Cross-language – Query in one language, docs in other
    - English, French, Spanish, Italian, German, Chinese, Arabic
  - Genomics
  - Spoken Document Retrieval

TREC Tracks

- Track: Basic task organization

- Previous tracks:
  - Ad-hoc – Basic retrieval from fixed document set
  - Cross-language – Query in one language, docs in other
    - English, French, Spanish, Italian, German, Chinese, Arabic
  - Genomics
  - Spoken Document Retrieval
  - Video search

TREC Tracks

- Track: Basic task organization

- Previous tracks:
  - Ad-hoc – Basic retrieval from fixed document set
  - Cross-language – Query in one language, docs in other
    - English, French, Spanish, Italian, German, Chinese, Arabic
  - Genomics
  - Spoken Document Retrieval
  - Video search
  - Question Answering

Current TREC tracks

- TREC 2014:
  - Contextual Suggestion
  - Clinical Decision Support Track
  - Federated Web Search
  - Knowledge-base Acceleration
  - Microblog
  - Session
  - Temporal Summarization
  - Web

Other Shared Tasks

- International:
  - CLEF (Europe); NTCIR (Japan); FIRE (India)

Other Shared Tasks

- International:
  - CLEF (Europe); NTCIR (Japan); FIRE (India)

- Other NIST:
  - DUC (Document Summarization)
  - Machine Translation
  - Topic Detection & Tracking

Other Shared Tasks

- International:
  - CLEF (Europe); NTCIR (Japan); FIRE (India)

- Other NIST:
  - DUC (Document Summarization)
  - Machine Translation
  - Topic Detection & Tracking

- Various:
  - CoNLL (NE, parsing,..); SENSEVAL: WSD; PASCAL (morphology); BioNLP (biological entities, relations)

Other Shared Tasks

- International:
  - CLEF (Europe); NTCIR (Japan); FIRE (India)

- Other NIST:
  - DUC (Document Summarization)
  - Machine Translation
  - Topic Detection & Tracking

- Various:
  - CoNLL (NE, parsing,..); SENSEVAL: WSD; PASCAL (morphology); BioNLP (biological entities, relations)
  - Mediaeval (multi-media information access)

TREC Question-Answering

- Several years (1999-2007)
  - Started with pure factoid questions from news sources

TREC Question-Answering

- Several years (1999-2007)
  - Started with pure factoid questions from news sources
  - Extended to lists, relationship

TREC Question-Answering

- Several years (1999-2007)
  - Started with pure factoid questions from news sources
  - Extended to lists, relationship
  - Extended to blog data
  - Employed question series

TREC Question-Answering

- Several years (1999-2007)
  - Started with pure factoid questions from news sources
  - Extended to lists, relationship
  - Extended to blog data
  - Employed question series
  - Added temporal constraints
  - 'Complex, interactive' evaluation

TREC Question-Answering

- Several years (1999-2007)
  - Started with pure factoid questions from news sources
  - Extended to lists, relationship
  - Extended to blog data
  - Employed question series
  - Added temporal constraints
  - 'Complex, interactive' evaluation

- Combined with summarization to form TAC (2008---)
  - Text Analytics Conference
    - Opinion Q/A, Knowledge-based population, Scientific Summarization

TREC Question-Answering

- Provides:
  - Lists of questions
  - Document collections (licensed via LDC)
  - Ranked document results
  - Evaluation tools: Answer verification patterns
  - Derived resources:
    - E.g. Roth and Li's question categories, training/test
  - Reams of related publications

Questions

- <top>
  - <num> Number: 894
  - <desc> Description: How far is it from Denver to Aspen?

- </top>

Questions

- <top>
  - <num> Number: 894
  - <desc> Description: How far is it from Denver to Aspen?

- </top>

- <top>
  - <num> Number: 895
  - <desc> Description: What county is Modesto, California in?

- </top>

Documents

- <DOC><DOCNO> APW20000817.0002 </DOCNO>

- <DOCTYPE> NEWS STORY </DOCTYPE><DATE_TIME> 2000-08-17 00:05 </DATE_TIME>

- <BODY> <HEADLINE> 19 charged with drug trafficking </HEADLINE>

- <TEXT><P>

- UTICA, N.Y. (AP) - Nineteen people involved in a drug trafficking ring in the Utica area were arrested early Wednesday, police said.

- </P><P>

- Those arrested are linked to 22 others picked up in May and comprise "a major cocaine, crack cocaine and marijuana distribution organization," according to the U.S. Department of Justice.

- </P>

Answer Keys

- 1394: French

- 1395: Nicole Kidman

- 1396: Vesuvius

- 1397: 62,046

- 1398: 1867

- 1399: Brigadoon

Reminder

- Team up!