NLP Applications

Two main areas:

• Massive management of textual information sources:
  • For human use
  • For automatic collection of linguistic resources
• Person/Machine interaction
NLP Applications

Massive management of textual information sources

- Machine Translation (MT)
- Information Retrieval (IR)
- Question Answering (Q&A)
- Information Extraction (IE)
- Summarization
Machine Translation

• Process of translating a text from a source language to a target language preserving some properties
  • The main property to preserve (but not the only one) is the meaning
• MT textual vs oral
• Different degrees of human intervention
Machine Translation

- Human Translation with Machine Support
- Machine Translation with Human Support
- Fully Automated Translation
  - Empirical systems
  - Rule-based systems
  - Statistical Machine Translation
  - Example-based Translation
Machine Translation

Interlingua

Semantic Str. → Semantic Str.

Semantic Transfer

Syntactic Str. → Syntactic Str.

Syntactic Transfer

Lexic Structure → Lexic Structure

Direct translation

Source text → Target text
Statistical Machine Translation

Translation Model $P(f|e)$

- Model for each word in the source language:
  - Its translation
  - the number of necessary words in the target language
  - the position of the translation in the sentence
  - the number of words that need to be generated from scratch.
NULL Quan tornes a casa?

NULL Quan tornes tornes tornes tornes casa?

NULL When are coming back home?

you When are coming back home?

When are you coming back home?
Information Retrieval

- Input
  - A collection of documents
    - The Web
    - A corporate document collection
    - ...
  - A user need represented as a query
- Output
  - The documents of the collection that satisfy the user needs.
Information Retrieval

Queries space: $Q$

Documents space: $D$

Representation space: $R$

Human judgement: $j$

Comparison function: $c$

Oard, 1997
Information Retrieval

IR types

• Type of information
  • Text, speech, structured information
• Query language
  • Exact, ambiguous
• Matching
  • Exact, approximate
• Kind of information needed
  • Loose, precise
• Relevance:
  • Usefulness of information according to user needs
• Natural extension of IR
• A QA system receives a query expressed in NL and tries to provide not a document containing the answer but the proper answer (usually a fact).
• QA systems need to use NLP techniques for both processing the question and looking for the answer.
Some QA systems that can be accessed through the Web:

- Webclopedia
- AskJeeves
  - http://www.ask.com
- LCC
  - http://www.languagecomputer.com/
Question Answering

- Factual QA
  - Who? When? Where?
- List QA
  - Which are the last 10 presidents of USA?
- Domain independent vs domain restricted QA
- QA with complex queries:
  - Which are the USA republican presidents after world war II?
- Linked queries
Question Answering

Frequently performed sequentially

Question Processing → Relevant terms
              Question type
              Focus

IR of relevant documents → Relevant documents

Segmentation in passages, IR of relevant passages → Relevant passages

Answer Extraction → answer
A summary is a reductive transformation of a source text into a summary text by extraction or generation.
Automatic Summarization

• Look for the relevant parts of a document and produce a summary of them
• Summarization vs Information Extraction
  - Information Extraction
    • What has to be extracted is defined a priori
      • “I am interested on this, look for it”
  - Summarization
    • An a priori definition of what is relevant is not always defined
Automatic Summarization

- Basic schema
  - multi-document
  - single-document
- query
- restrictions
  - extract
  - abstract
  - headline
- Summarizer
Information Extraction

- Extracting useful information from free text
- MUC, ACE, TAC challenges
- Named Entity Recognition (NER)
- Named Entity Classification (NEC)
- Both tasks together (NERC)
- Slot Filling
- Relation Extraction
Information Extraction

NERC

y  B-PER  O  B-QNT  O  O  B-ORG  I-ORG  
  Jim  bought  300  shares  of  Acme  Corp.

y  B-PER  I-PER  O  O  B-LOC  
  Jack  London  went  to  Paris

y  B-PER  I-PER  O  O  B-LOC  
  Paris  Hilton  went  to  London
Information Extraction

Slot Filling

- Set of relevant slots
- ML
  - Supervised Learning
  - Unsupervised Learning
    - Distant learning
  - Semisupervised Learning
    - Active Learning
- Rule-based systems
Human/machine communication

Main goal
Help users perform specific tasks according to their objectives

Tasks of the dialogue systems
- Interpreting user intervention
- Dialogue Management
- Generating system's intervention
Using the Natural Language Mode

• Advantages
  • Human Language (natural, friendly)
  • Complex ideas can be expressed
  • References to other entities are easy to express

• Disadvantages
  • High cost
  • Ambiguity -- mistakes
  • Limitations for accessing several applications (such as graphics)

• Appropriate for occasional access to applications that need to express complex operations (especially when domain can be restricted)
An Example of Conversational System

System: Welcome to the information service, what do you want?
User: I want to go from Barcelona to Valencia
System: When do you want to go?
User: Next Tuesday
System: At what time, morning or afternoon?
User: Morning, please
System: There are 3 Euromed trains on Tuesday morning
Tasks of the Dialogue Systems

• Interpreting user’s intervention
  – Using dialog and domain knowledge

• Dialogue Management
  – Determine next system actions considering user's intention

• Answer Generation
  – Generate the system's sentences most appropriate at each state of the dialogue
Interpreting the user intervention

• Goal: understanding user's intention
• Knowledge involved
  • Phonetics and phonology
  • Morphology
  • Syntax
  • Semantics (lexical and compositional)
  • Pragmatics
  • Discourse
Interpreting the user intervention

- Goal: understanding the user's intention
- Precise information from the user is required
- The complexity of this process depends on the system
  - Complete (deep) syntactic and semantic analysis
  - Partial (shallow) syntactic and semantic analysis
  - Processing key words
- This process is restricted by considering limited applications tasks
Intention Recognition (Real systems)

- The system infers the application task the user is asking for
  - Application: Giving information on cultural events
    - *Time or place where a specific event takes place*
    - *Events that take place in a specific place*
  - Application: Giving information on trains
    - *Schedule for a specific train*
- The system asks the user the information the application needs
  - The system ignores the information not useful for the application
Intention Recognition (Real systems)

- System initiative
- User initiative very limited
  - Not allowed in complex acts such as confirmation, clarification and indirect answers

S1: Which is your account number?
U1: My account number in Online Bank?

S2: Would you want to transfer 1500 euros to your new account?
U2: If I have this amount, ok
Representation of a user intervention

Reservation
  What
  Quantity

Train ticket
  Date
  Type
  From
  To
  Hour
  Prize

What
  Quantity: 1

Type: Euromed
  From: Barcelona
  To: Valencia
Dialogue Management

• Controlling dialog to help the user to achieve his goals
  – At each step of the conversation
    • **Who can speak**
    • **What can be said**

  – **Used information**
    • Interpretation of the user intervention
    • Application (domain) knowledge
Dialogue Management

• Determine the next system's action(s)
  – Answer user's questions
  – Ask the user for more information
  – Confirm/Clarify user's interventions
  – Notify problems when accessing the application
  – Suggest alternatives

• Generation of the system's messages
  – The content
  – The presentation
Dialogue Management

- Determine the next system's action(s)
  - Answer user's questions
  - Ask the user for more information
  - Confirm/Clarify user's interventions
  - Notify problems when accessing the application
  - Suggest alternatives

- Generation of the system's messages
  - The content
  - The presentation
Dialogue Management

• Determine the next system's action(s)
  – Answer user's questions
  – Ask the user for more information
  – Confirm/Clarify user's interventions
  – Notify problems when accessing the application
  – Suggest alternatives

• Generation of the system's messages
  – The content
  – The presentation
Dialogue Management

• **Research systems**
  • Focused on the development of models and algorithms for supporting several dialogue phenomena for complex tasks

• **Real systems**
  • Focused on the development of robust strategies, to deal efficiently with most common dialogue phenomena for simple applications
Answer Generation

• Generation of sentences to achieve the goals the dialogue manager has selected

• Tasks
  • Content selection
    • Presentation of content
    • Using rethorical elements
  • Superficial realization
    • Semantic representation of the text
    • What to say, how to say
Content Selection

• Determine the content of the system sentences in order to achieve the goals

• Examples:
  • Madagascar is not shown in Sant Cugat [Nucleus]
    • It is shown in Barcelona [Satellite]
  • Would you like a suite? [Nucleus]
    • It is the same price than the doble room [Satellite]
  • Magic Flaute is not shown this year at Liceu [Nucleus]
    • But Figaro Wedding is [Satellite]
Superficial realization

• Goal: to determine **how** content selected is presented

• Examples:
  
  *Madagascar is not shown in Sant Cugat but it is shown in Barcelona city*

• Tasks
  
  • Construction of phrases
  • Lexical selection
Chatbots - “A service, powered by rules and sometimes artificial intelligence, that you interact with via a chat interface.

The service could be any number of things, ranging from functional to fun, and it could live in any major chat product (Facebook Messenger, Slack, Telegram, Text Messages, etc.).

https://chatbotsmagazine.com/
Chatbots


- Based on AIML: Artificial Intelligence Markup Language, based on XML.

- Facilitate the creation of virtual personal assistant apps (like Siri)
Language Generation

• Content planning
  • Semantic representation of the text
  • What to say, how to say

• Form planning
  • Presentation of content
  • Using rethorical elements
Automatic collection of linguistic resources

• Aligned corpora (various levels)
• Grammars
• Gazetteers
• Resources including
  – Morphology bases
  – Selectional restrictions
  – Subcategorization patterns
  – Topic Signatures