Chatbots and NLP

Damir Cavar

Indiana University

June 2018

Agenda

- Chatbots then
- and now
- NLP and Chatbots
- Issues
- Modest Prediction of Sensible R&D Directions

From Verbmobil to Chatbots

- Speech2Speech MT
  - Dialog Models
  - ASR
  - Machine Translation
  - NLP

- Chatbots and Avatars
  - Text input
  - Text to speech output
  - Virtual communication partner with lip-synchronization

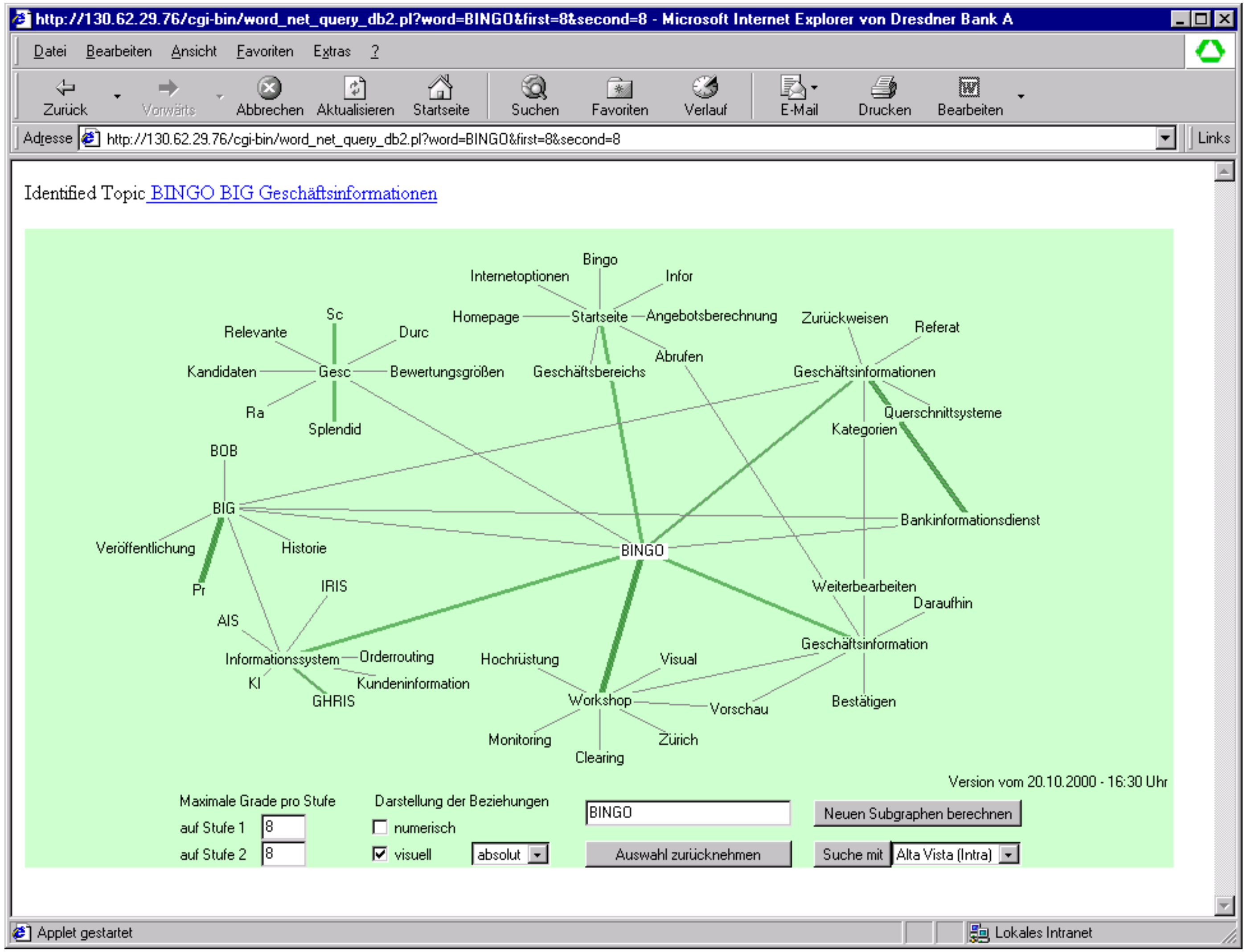
Avatar

- Web-based
- Terminals in stores

Valerie (Charamel GmbH)

Knowledge Graphs

- RDB-based SemNet
  - (before OWL, etc.)



NLP Technologies

- Back in 2000
  - Regular expressions and pattern matching
  - Template-based text generation
  - Finite State Dialog modeling
  - Knowledge Graphs (SemNets) on RDBs
  - Text2Speech

  - Part-of-speech tagging
  - Parsing
  - Machine Translation

NLP Technologies

- Focus on limited models and technologies:
  - Dependency Parse Trees
  - Treebank-derived Constituent Tree Parsers
  - Label/Tag-based Semantic Role Labeling
  - …
  - Pipeline-architecture as such:
    - Isolated modules with very limited NLP-focus chained in an input-output pipeline
      - CoreNLP, spaCy, OpenNLP, NLTK, UIMA, …

NLP Technologies

- State of the Art: (Sebastian Ruder’s overview)
  - Part-of-Speech Tagging:
    - Use: word-level part of speech annotation with a limited set of tags that encode some morphosyntactic features
    - **F1 score**: **95% - 97%** based on WSJ portion of Penn Treebank, more than 100 treebanks for UD
    - Best performing: Deep Learning Approaches (alternatives not evaluated)

NLP Technologies

- State of the Art: (Sebastian Ruder's overview)
  - Constituent Tree Parsing:
    - Use: phrasal structure; relations, hierarchies and ambiguities between phrases; semantic scope relation; …
    - **F1 score**: **92% - 95%** based on Penn Treebank
    - Best performing: Deep Learning Approaches (alternatives not evaluated)
  - Dependency Parsing:
    - Use: dependency relations between elements in the sentence; simplified annotation of functional relations: Subject, Object, Modifier, …
    - **F1 score** on labels and relations: **91% - 94%** based on Stanford Dependency conversion of the Penn Treebank
    - Best performing: Deep Learning Approaches (alternatives not evaluated)

NLP Technologies

- State of the Art: (Sebastian Ruder’s overview)
    - Named Entity Recognition:
        - Use: entity labeling – person, institution, location, time, currency, …
        - **F1 score**: **90% - 92%** based on Reuters RCV1 corpus with **four** NE-types (PER, LOC, ORG, MISC) using BIO notation
        - Best performing: Deep Learning Approaches (alternatives not evaluated)
    - Semantic Role Labeling:
        - Use: Label predicate argument structure (*Who gave what to who*): Predicate, Subject, Object, entity and relation extraction
        - F1 score: **81% - 84%** based on OntoNotes benchmark of the Penn Treebank
        - Best performing: Deep Learning Approaches (alternatives not evaluated)

NLP Technologies

- F1 score margins and error rates:
  - Basic token-level classification: error of approx. **4%**
  - Word-level annotation, syntactic parsing: **10%**
  - Semantic-level annotation: **30%**

- Not much has changed since 2000!

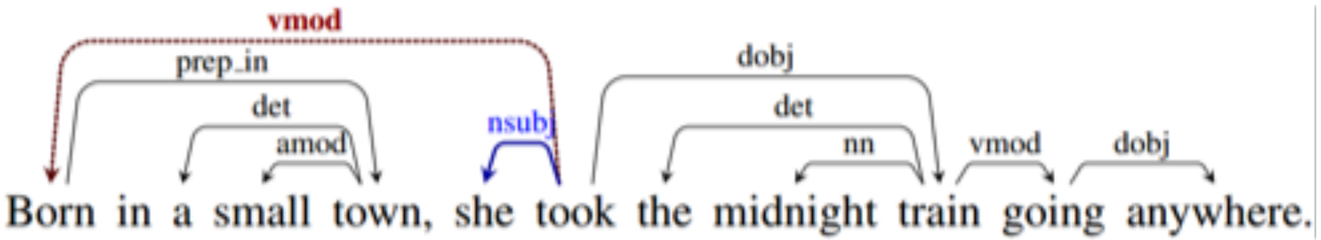
- Situation check:
  - Mono-culture of training/test-datasets for data driven ML/DL-methods
  - Limitation to weak linguistic models (e.g. *Constituent Trees*, *NE-classes*, *Semantic roles*), annotation standards (e.g. *Dependencies*)

NLP Technologies

- Situation check:
  - Limited use of NLP-pipelines: PoS-tagging, Lemmatization
    - CoreNLP: Constituent Parser; Dependency Parser; Coreference Analysis; …
    - spaCy: Dependency Parser
    - NLTK: WordNet
  - Lack of APIs that interface to linguistic output data structures
    - NLP developers lack understanding of the linguistic annotations generated by pipelines or tools

NLP Example

- Stanford Open IE (paper and website)



  - Lack of intuition of dependency relations
    - Modification of ROOT (took) by “born in a small town” is counterintuitive
- Lack of:
  - Clause level hierarchical relation analysis (subordinate clauses and scope)
  - Tempus, Mood, … annotation
  - Pragmatic and semantic properties (and relevant linguistic features)

NLP Example

- Scope between clauses:
  - Reuters reported [that [Google bought Apple]]
  - Reuters did not report [that [Google bought Apple]]
  - Reuters did not deny [that [Google bought Apple]]
- Tense:
  - Tim Cook bought Google.
  - Tim Cook will buy Google one day.

NLP Technologies

- Applied to real text:
  - Sentence length over 10 to 15 tokens breaks most probabilistic or NN parsers (Dependency parsers, in particular)
- Problematic domains, for example:
  - SEC, Financial, or Business Reports
  - Case-law and legal documents
  - Medical text (patient reports, documentations)
- Current free and open NLP-pipelines are of limited use.
- Are they of any use for Chatbot technologies?

State of the Art

- Δ between 2000 – 2018
  - ASR improvements
  - Knowledge Graphs, Ontologies
  - Integration
    - Data sources
    - Interfaces, multi-modal interaction
    - Device architecture
- Is there any significant progress in ___ ?
  - Dialog management
  - NLP at the utterance and discourse level
  - Semantics and Pragmatics

Knowledge Representations

- General World Knowledge
  - From static to dynamic, with inferencing, reasoning
- Domain Specific Knowledge
  - Medical, Financial, Business, Legal, etc.
- Discourse specific Knowledge
  - Simple dialog memory (concepts and their linguistic features, relevant for anaphora resolution, coreference analysis)
  - Knowledge Graph or Ontology of semantic concept space in encapsulated discourse

Directions

- Speech
  - Prosodie – Semantics / Pragmatics interface

- Semantics and Pragmatics
  - Entailment
  - Quantifiers
  - Scope
  - Implicatures
  - Presuppositions

Prosodie – Semantics / Pragmatics

- Suprasegmental speech properties
  - Intonation Contour
    - ASR output: what did you buy
      - Speech properties:
        - interrogative or rhetorical, echo question
  - Pitch Accent (in specific languages)
    - Simple accent detection:
      - White board vs. whiteboard
      - HOtel vs. hoTEL
  - Focus Stress:
    - Contrastive Stress: I bought THAT CAR. I BOUGHT that car.
    - Verum Focus: I DID buy that car.

Prosodie – Semantics / Pragmatics

- Data sources
  - Limited corpora for a few languages
    - Richer documentation without audio data component
  - Creation of corpora with relevant speech properties and linguistic annotation (mapping of content to some semantic and pragmatic level)
- Efficient speech signal level processing:
  - librosa
  - OpenSMILE
  - Etc.

Speech Acts, Implicatures, Presuppositions

- Deep Linguistic Processing:
  - A to B: “I bought the blue car.”
  - Implicature:
    - A and B talked about the event earlier.
    - There is a set of cars, at least 2 that was in the range of A’s action.
    - None of the other cars in the set is blue.
  - Linguistic indicators:
    - Definiteness via “the”
    - Specificity of the Noun Phrase

Speech Acts, Implicatures, Presuppositions

- Deep Linguistic Processing:
  - "Peter fed his cat."
  - Presupposition:
    - Peter owns a cat.
    - Peter owns cat food.
    - …
  - Linguistic indicators:
    - Possessive
- Types:
  - Universal linguistic properties (see Grice Maxims, Relevance Theory)
  - Language specific properties (dependency to cultural and sociological aspects)
  - Domain specific: e.g. "to be like milk"

NLP and Future Chatbots

- Technology trends might include:
  - Speech signal properties that relate to:
    - Sentiment, Emotions
    - Prosodie and Semantics/Pragmatics Interface
    - Properties of speakers, general and temporary
  - Pragmatics and Semantics
    - Implicatures and Presuppositions, Natural Logic, Entailment
    - Irony, Humor, Sarcasm, Deception…
  - Knowledge Representation
    - Dynamic, Cascaded or Interconnected KRs
  - Multi-Modal systems (speech/language, image, gesture, mimic)
  - NLP
    - Deep Learning and hybrid systems, knowledge-based engineering and data-driven (M|D)L