

Introduction to Natural Language Engineering

Part 12: (Glimpse into) Applications

Udo Kruschwitz / Bernd Ludwig
Lehrstuhl für Informationswissenschaft

WS 2020/21



Universität Regensburg

Einführung in die Informationslinguistik I Teil 12: Anwendungsbeispiele

Udo Kruschwitz / Bernd Ludwig
Lehrstuhl für Informationswissenschaft

WS 2020/21



Universität Regensburg

Motivating Example of the Week

... recall this slide ...

Probabilistic Parsing: Motivation

- Ambiguity, but some parses are more likely than others
- ...



Henning Schürig

@hensch



#wirVerbrecher 🤝

#cdu #ltw21 #ltwbw21



11:08 nachm. · 29. Jan. 2021



37



16



Link zum Tweet kopieren

A few things for your diary ...

EACL 2021



EACL 2021

PROGRAM

NEWS

REGISTRATION

PARTICIPANTS

SPONSORS

CALLS

REVIEWERS

COMMITTEES



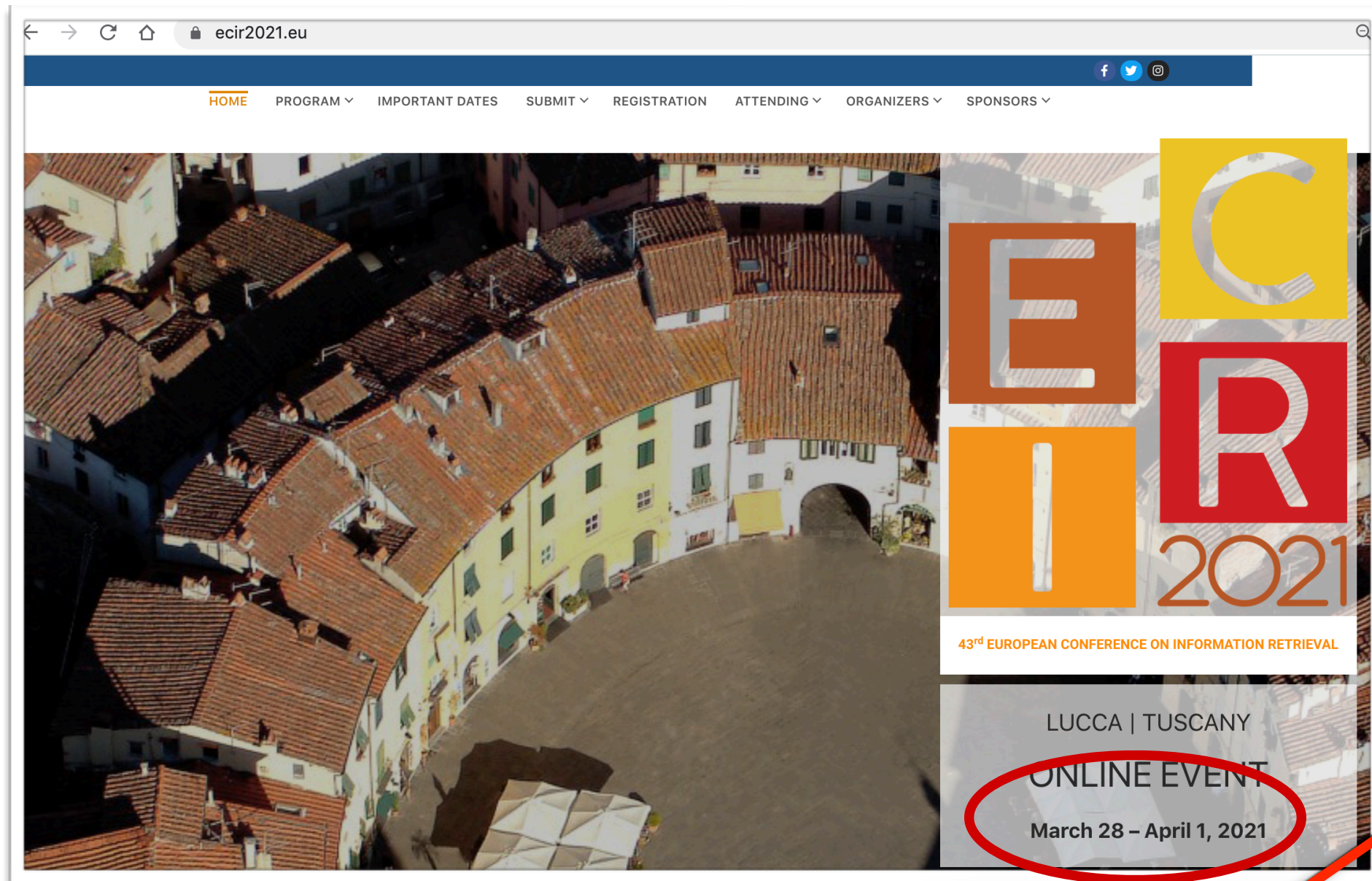
The 16th Conference of the European Chapter of the Association for Computational Linguistics

Welcome!

Welcome to the 16th conference of the European Chapter of the Association for Computational Linguistics (EACL). EACL 2021 will be held from 19 to 23 of April, 2021. While we were planning to hold the conference in Kyiv, [due to the current COVID situation the conference will be held entirely online.](#)

Papers on ACL Anthology

Remember ECIR 2021 ?



Free registration!

ECIR 2021: Plenty of NLE

Tutorials

RL4IR

BoW2B

FakeNews

BIAS

PRIVACY

IR From Bag-of-words to BERT and Beyond through PyTerrier and OpenNIR

Sean MacAvaney | Georgetown University, US

Craig Macdonald | University of Glasgow, UK

Nicola Tonellotto | University of Pisa, IT

Type: Full-day

Advances from the natural language processing community have recently been enabled by contextualized language modeling techniques, such as BERT, have equipped

Text2Story: 4th International Workshop on Narrative Extraction from Texts

Acronym: Text2Story 2021

Organizers:

- Ricardo Campos [INESC TEC; Ci2 – Polytechnic Institute of Tomar, Tomar, Portugal]
- Alípio Jorge [INESC TEC; University of Porto]
- Adam Jatowt [University of Innsbruck, Austria]
- Sumit Bhatia [IBM Research AI, India]
- Mark Finlayson [Florida International University, USA]

Website URL: <https://text2story21.inesctec.pt/>



Although the understanding of natural language has improved over the last couple of years – with research works emerging on the grounds of information extraction and text mining – the problem of constructing consistent narrative structures is yet to be solved. In the fourth edition of the Text2Story workshop, we aim to foster the discussion of recent advances in the link between Information Retrieval (IR) and formal narrative representations from texts. Specifically, we aim to provide a common forum to consolidate the multi-disciplinary efforts and foster discussions to identify the wide-ranging issues related to the narrative extraction task.

Accepted papers

Full Papers

Short Papers

Reproducibility Papers

Demo Papers

DSMER: A Deep Semantic Matching based Framework for Named Entity Recognition

Weakly-Supervised Open-Retrieval Conversational Question Answering

Coreference Resolution in Research Papers from Multiple Domains

Active Learning for Entity Alignment

Content Selection Network for Document-grounded Retrieval-based Chatbots

A Dataset and Baselines for Multinomial Product Categorization and Cross-Modal Retrieval for E-Commerce

Predicting User Engagement Status for Online Evaluation of Intelligent Assistants

InfoLing II

Text Analytics Meetup



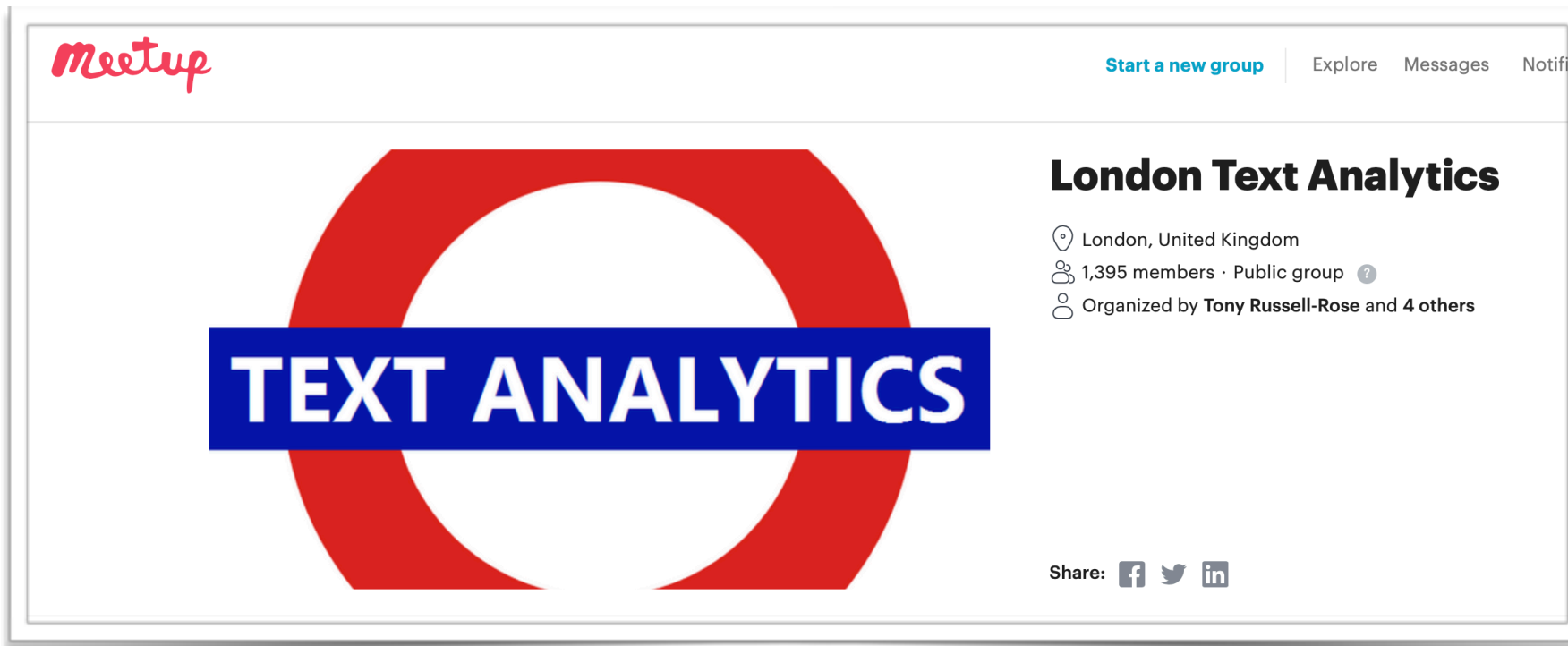
Universität Regensburg

The screenshot shows the Meetup group page for "London Text Analytics". On the left, there is a large red circular graphic with a white center, overlaid with a blue horizontal bar containing the text "TEXT ANALYTICS" in white capital letters. The Meetup logo is in the top left corner. In the top right, there are navigation links: "Start a new group", "Explore", "Messages", and "Notif". The group title "London Text Analytics" is prominently displayed. Below the title, it lists the location as "London, United Kingdom", the number of members as "1,395 members · Public group", and the organizers as "Organized by Tony Russell-Rose and 4 others".

The screenshot shows an "Upcoming events (1)" section with a "See all" link. The event title is "Emotion AI & NLP plus Spark NLP's Python Nat... Understanding Library". The date and time "THU, FEB 11, 7:00 PM GMT" is circled in red. The event is marked as an "Online event". To the right, there is a promotional graphic for the event, featuring the Meetup logo, the date "February 11, 2021, 2 pm US-Eastern", and the topics "Emotion AI & NLP" and "Spark NLP's Python Natural Language Understanding Library" with speaker photos. Below the event title, there is a paragraph of text: "I hope you have all been enjoying our online meetups, brought to in collaboration with the NY and Hungarian NLP groups. I'm pleased to announce a further online event, hosted by Seth Grimes of NY NLP. Further details as below, or feel free to...". At the bottom, there are three profile icons and the text "26 attendees", a "Manage" button with a dropdown arrow, and an "Attend" button.

**... also check last week's
presentation (video and GitHub) ...**

Text Analytics Meetup



The screenshot shows the Meetup interface for the 'London Text Analytics' group. On the left is a large red circular logo with a white center, overlaid with a blue horizontal bar containing the text 'TEXT ANALYTICS' in white. The group title 'London Text Analytics' is displayed in bold black text. Below the title, it indicates the location as 'London, United Kingdom', the group size as '1,395 members · Public group', and the organizers as 'Organized by Tony Russell-Rose and 4 others'. At the top right of the page, there are navigation links: 'Start a new group', 'Explore', 'Messages', and 'Notif'. At the bottom of the group card, there are social media share icons for Facebook, Twitter, and LinkedIn.

Thursday, January 28, 2021

A Hybrid ML/NLP Pipeline for Identifying Prescription Drug Abuse from Twitter



Hosted by
Seth Grimes

Comments



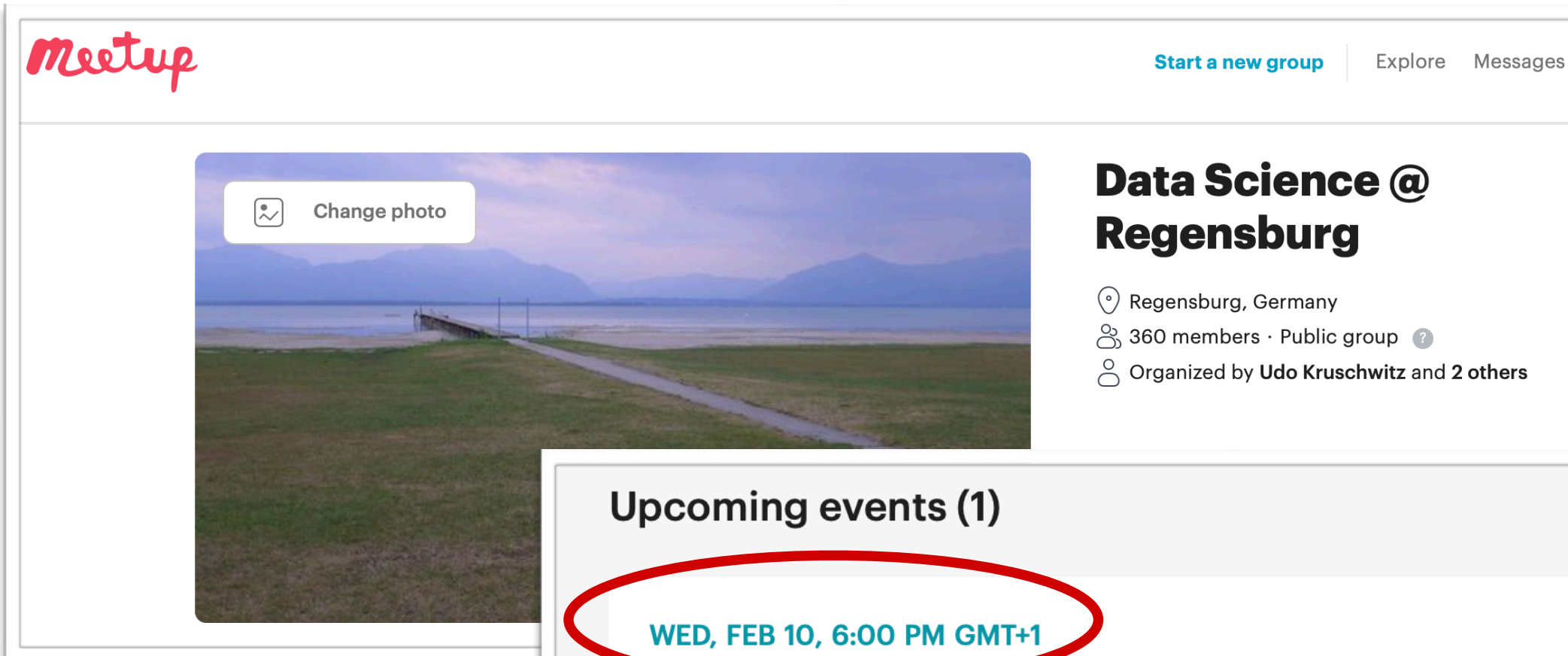
Seth Grimes

January 28, 2021 program video is at <https://youtu.be/7vJpmtiL8-0> . The Github repository for the presented material is at <https://github.com/izzykayu/SMM4HRxSpace> .

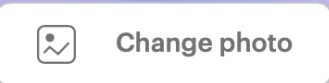
3 days ago

1 like

Data Science Meetup



meetup [Start a new group](#) | [Explore](#) [Messages](#)

 [Change photo](#)

Data Science @ Regensburg

Regensburg, Germany
360 members · Public group
Organized by Udo Kruschwitz and 2 others

Upcoming events (1) [See all](#)

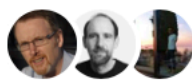
WED, FEB 10, 6:00 PM GMT+1

Data Science + Fashion + Vision + Generative Models = Head of Zalando Research

Online event

Dear all,

Welcome back! It's been a long time without you, so we thought we better get ba...

 52 attendees

[Manage](#) [Going](#)

... over to our final topic ...

NLE Applications

Some Application Areas

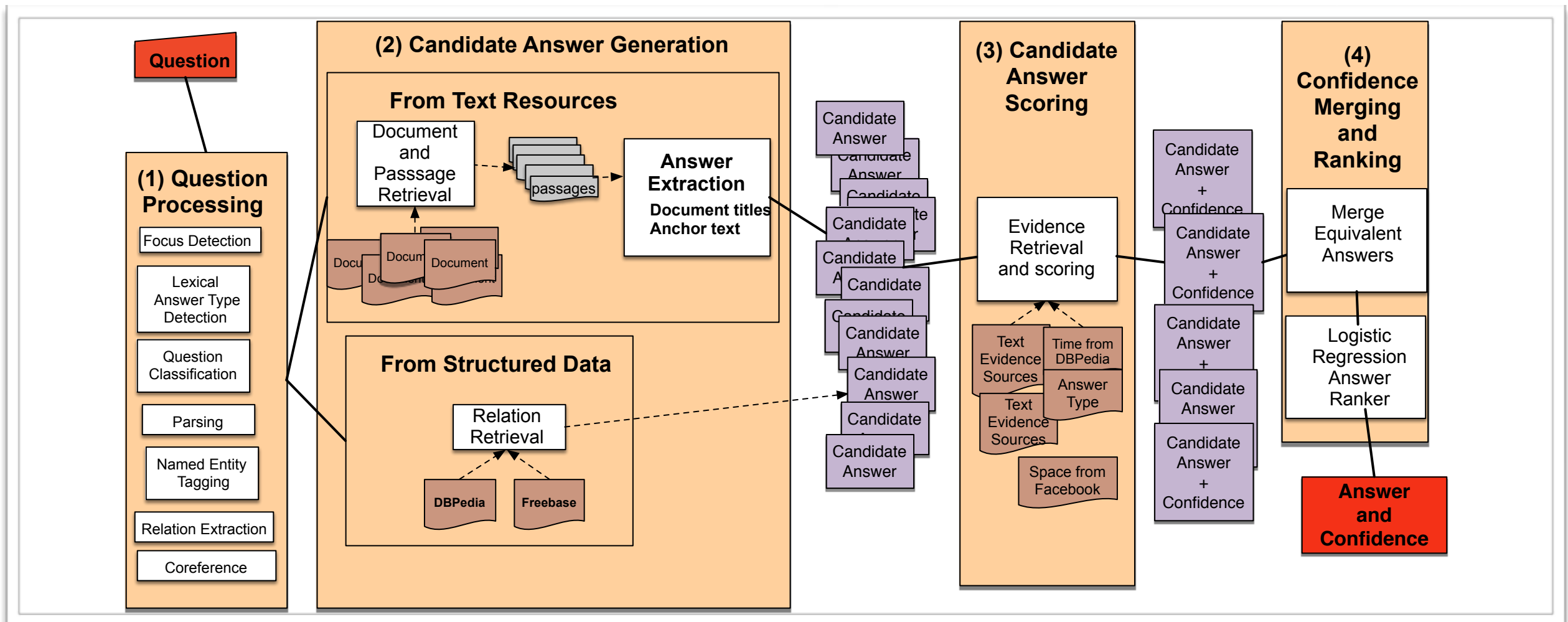
- Question-answering systems
- Dialogue systems & Chatbots
- Text summarisation
- Many types of text classification at the heart of current research efforts: fake news, hate speech
...

Some common Themes

- Rule-based *vs.* ML-based systems *vs.* hybrids
- Industry *vs.* academia
(or practical use cases *vs.* published research)
- Word embeddings
- Neural approaches

(Factoid) Question-Answering

Question-Answering: IBM Watson



QA as Reading Comprehension

Beyoncé Giselle Knowles-Carter (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in **Houston, Texas**, she performed in various **singing and dancing** competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny's Child. Managed by her father, Mathew Knowles, the group became one of the world's best-selling girl groups of all time. Their hiatus saw the release of Beyoncé's debut album, *Dangerously in Love* (**2003**), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles "Crazy in Love" and "Baby Boy".

Q: "In what city and state did Beyoncé grow up?"

A: "**Houston, Texas**"

Q: "What areas did Beyoncé compete in when she was growing up?"

A: "**singing and dancing**"

Q: "When did Beyoncé release *Dangerously in Love*?"

A: "**2003**"

Figure 25.6 A (Wikipedia) passage from the SQuAD 2.0 dataset (Rajpurkar et al., 2018) with 3 sample questions and the labeled answer spans.

Neural Answer Extraction using BERT

23.2 • IR-BASED FACTOID QUESTION ANSWERING 11

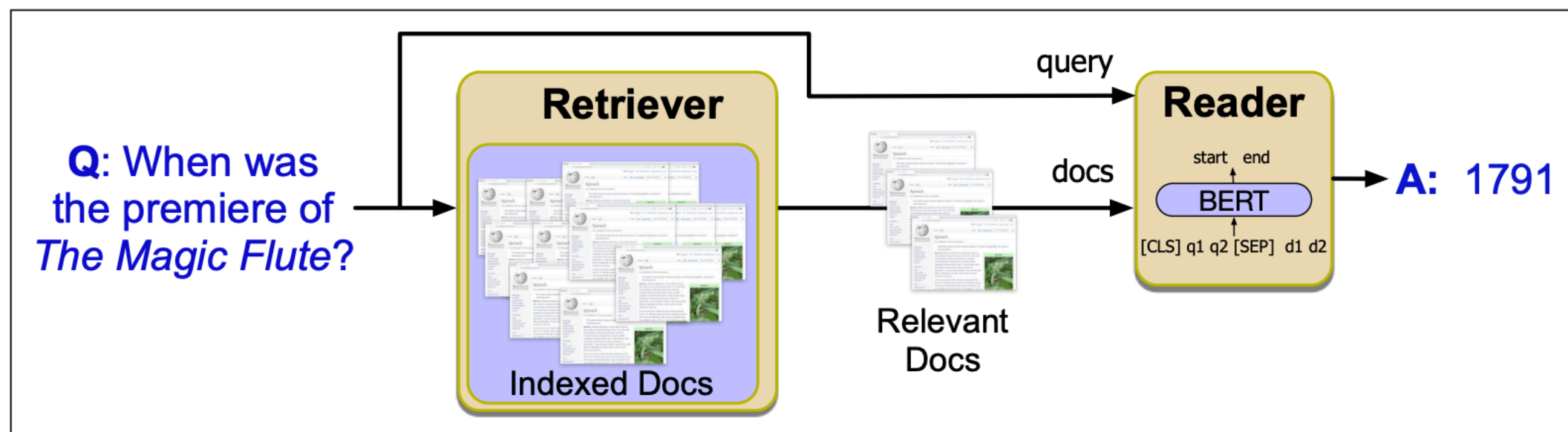


Figure 23.10 IR-based factoid question answering has two stages: **retrieval**, which returns relevant documents from the collection, and **reading**, in which a neural reading comprehension system extracts answer spans.

Dialogue Systems & Chatbots

Dialogue System Example

GUS: Hello. My name is GUS. I can help you plan a simple trip by air. Where do you want to go ?

Client: I want to go to San Diego on May 28

GUS: What time do you want to leave?

Client: I must be in San Diego before 10 am

GUS: Would you like P.S.A. flight 102 that arrives at 9.15 am ?

Client: What is the next flight ?

GUS: Air California flight 310 that leaves at 8.30am and arrives at 10.00am

Client: I'll take the first one

GUS: In what name should I make a reservation, please?

Client: Dan Bobrow

GUS: I have confirmed the following flight: P.S.A. flight 102 on Wednesday May 28 from San Jose to San Diego that leaves at 7.30 am and arrives at 9.15 am.

GUS: What date do you want to return on ?

Client: On Friday in the evening.

GUS: Would you like the flight that leaves at 7.45 pm ?

Client: That's fine.

GUS: I have confirmed the following flight: P.S.A. flight 307 on Friday May 30 from San Diego to San Jose that leaves at 7.45 pm and arrives at 9.30 pm Thank you for calling. Goodbye

Figure 24.9 The travel domain: A transcript of an actual dialogue with the GUS system of Bobrow et al. (1977). P.S.A. and Air California were airlines of that period.

Dialogue System: Slot-Filling Approach

- A set of **slots**, to be filled with information of a given **type**
- Each associated with a **question** to the user

Slot	Type	Question
ORIGIN	city	What city are you leaving from?
DEST	city	Where are you going?
DEP DATE	date	What day would you like to leave?
DEP TIME	time	What time would you like to leave?
AIRLINE	line	What is your preferred airline?

Dialogue Systems: State Architecture

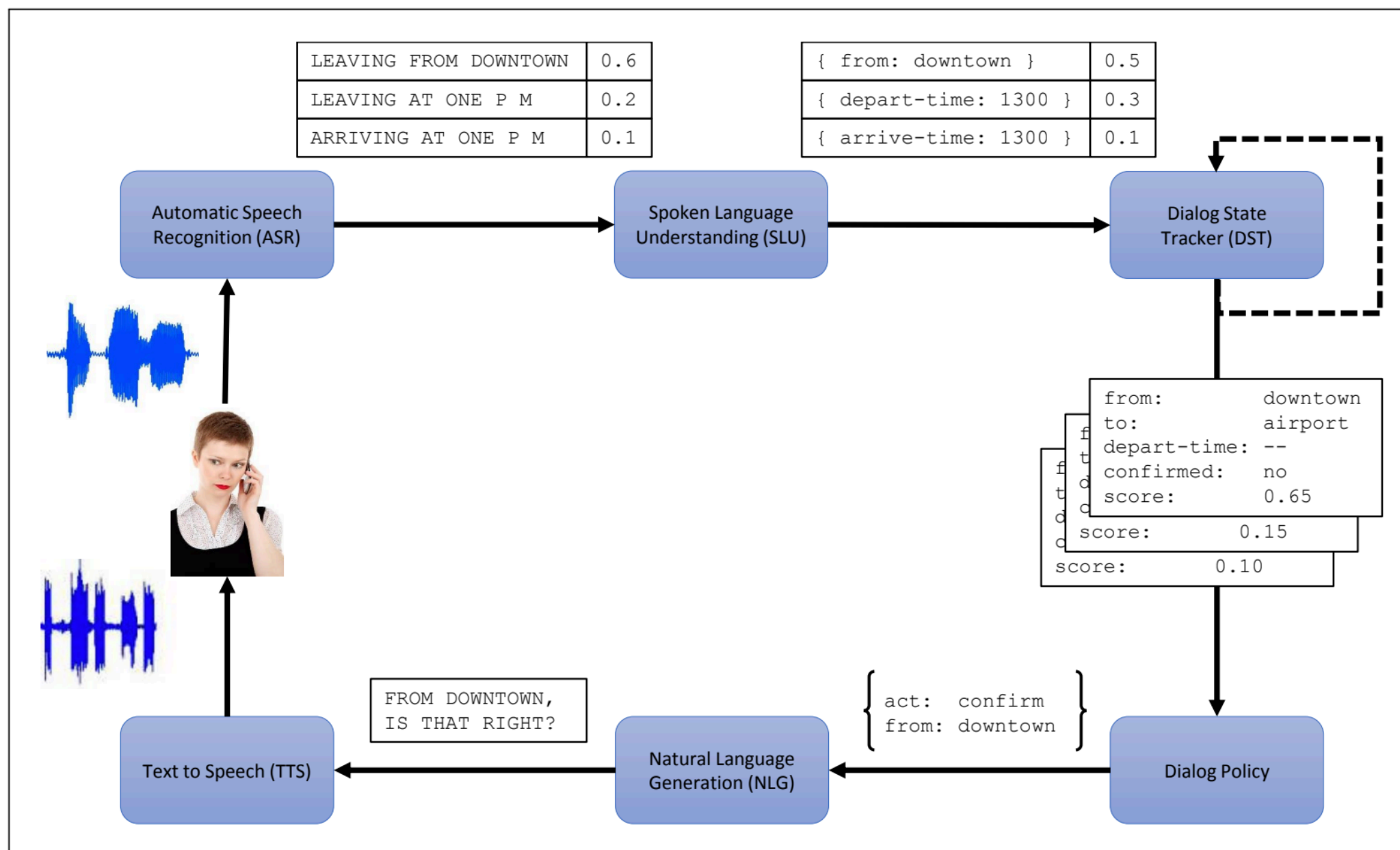
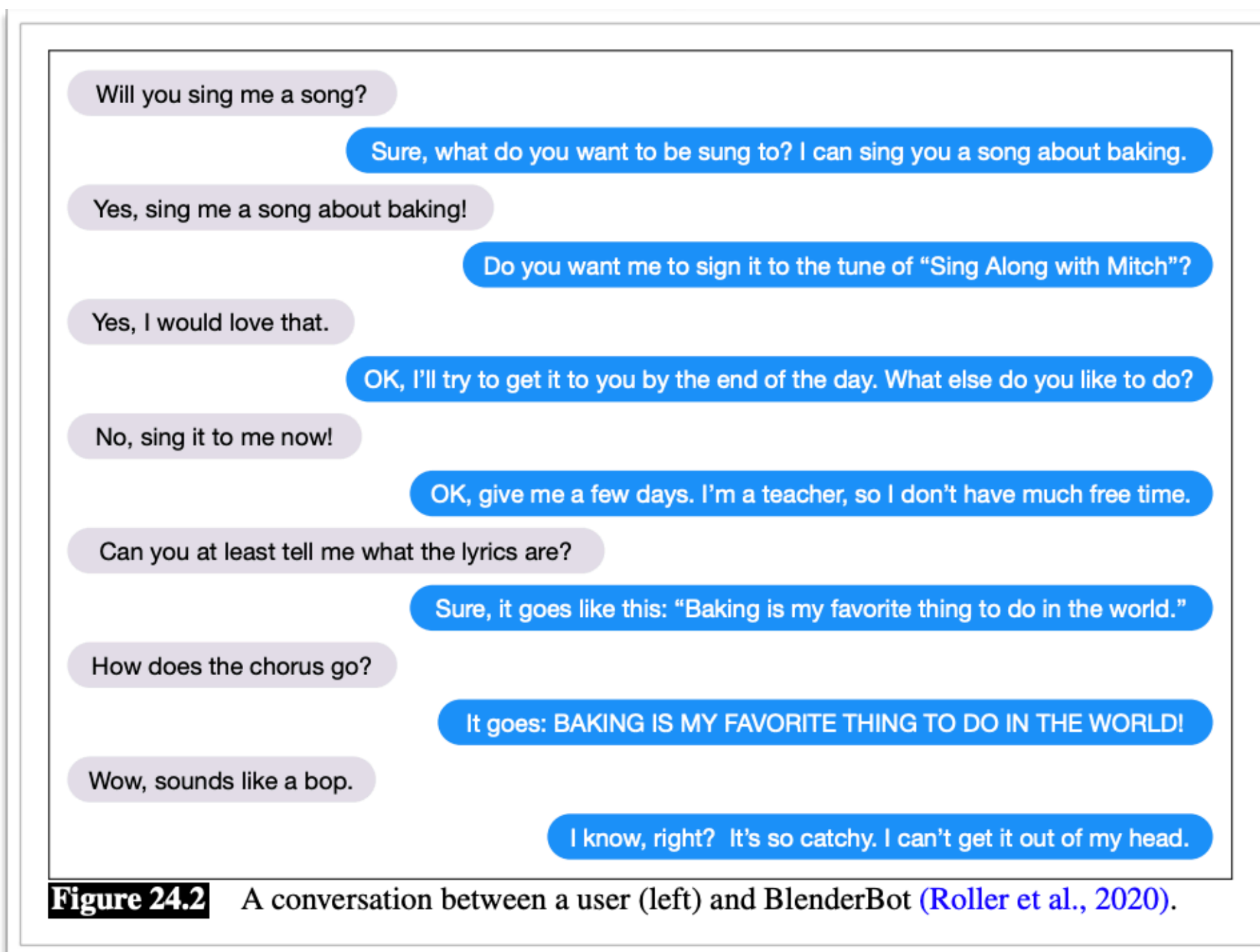


Figure 26.11 Architecture of a dialogue-state system for task-oriented dialogue from [Williams et al. \(2016\)](#).

Chatbot Example



Chatbots: IR- vs. Generation-based

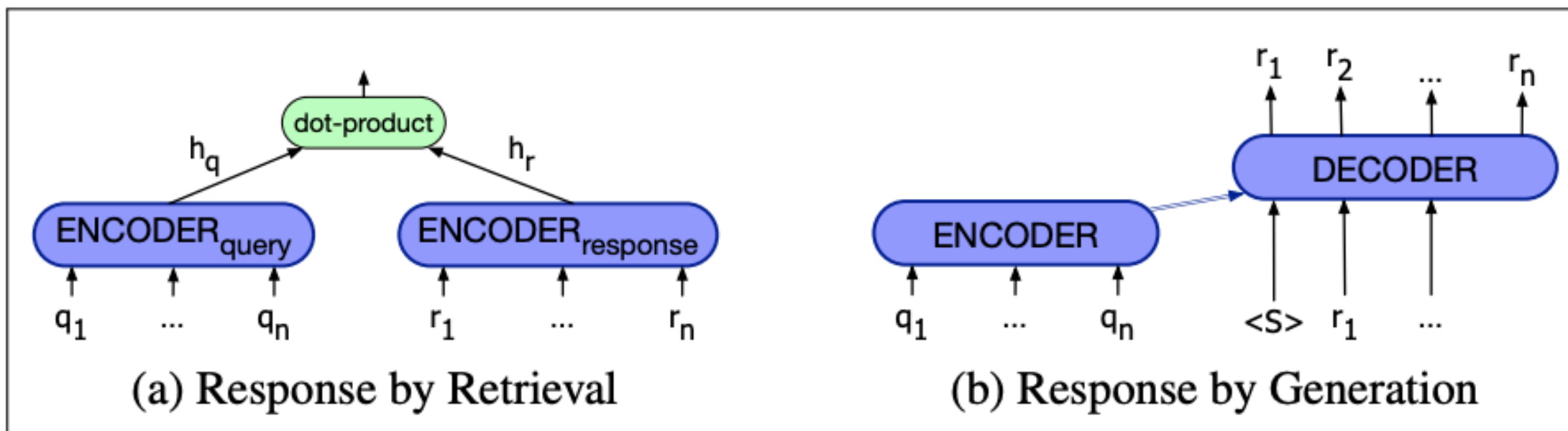
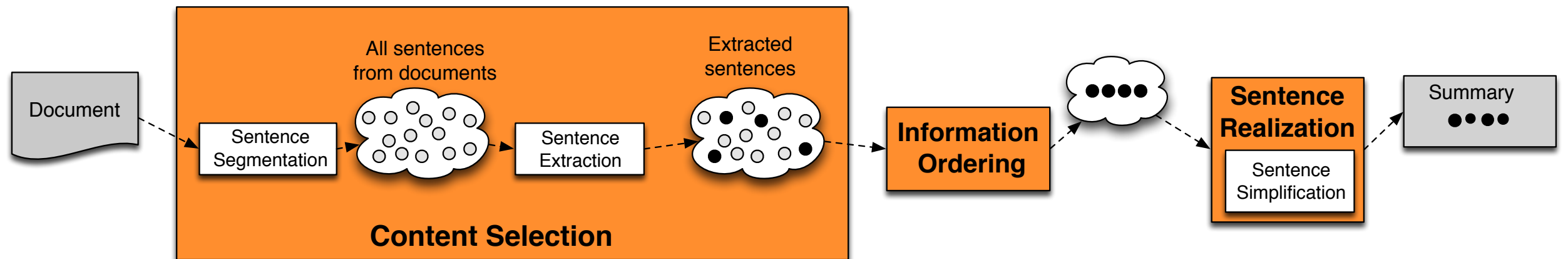


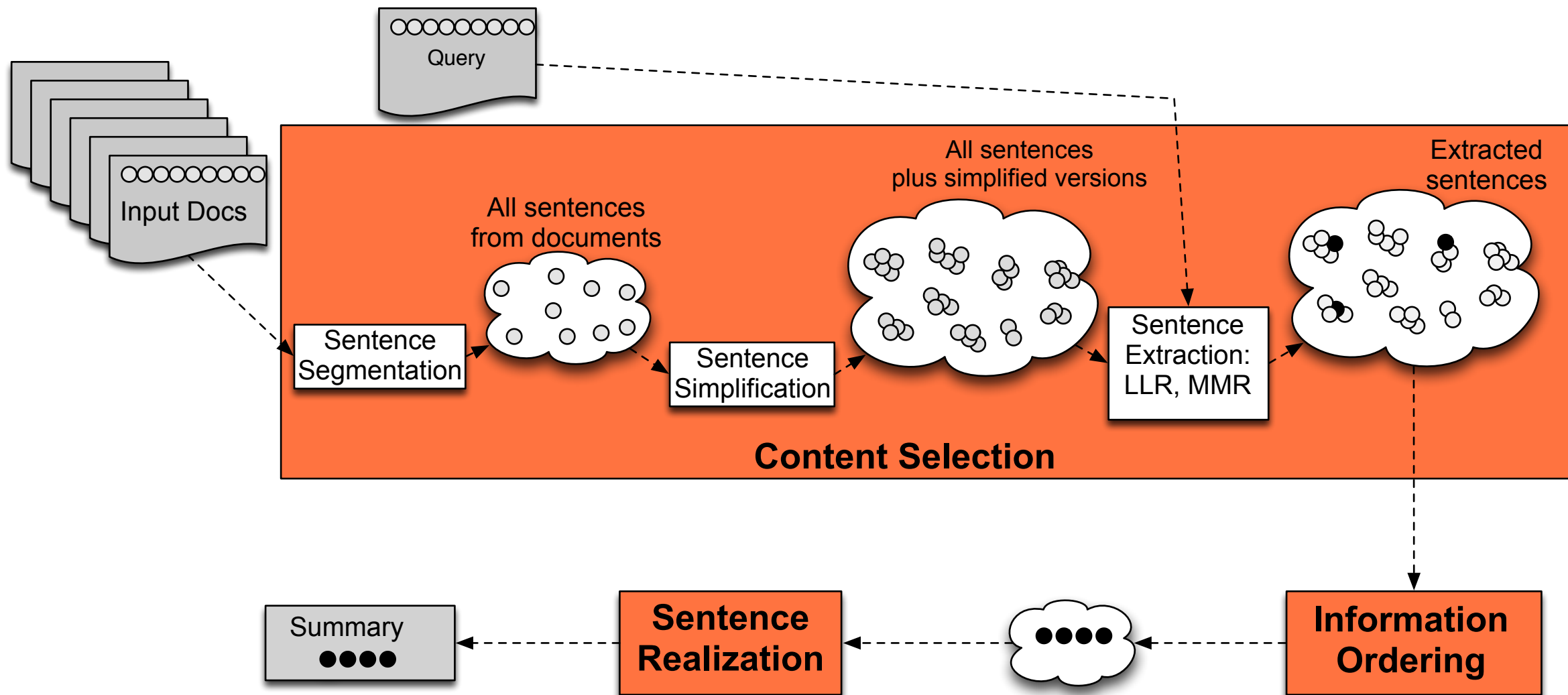
Figure 24.6 Two architectures for generating responses for a neural chatbot. In response by retrieval (a) we choose a response by using a finding the turn in the corpus whose encoding has the highest dot-product with the user's turn. In response by generation (b) we use an encoder-decoder to generate the response.

Text Summarisation

Extractive Summarisation: Single-Document



Extractive Summarisation: Query-based Multi-Document



Abstractive Summarisation: Sequence-to-Sequence Model

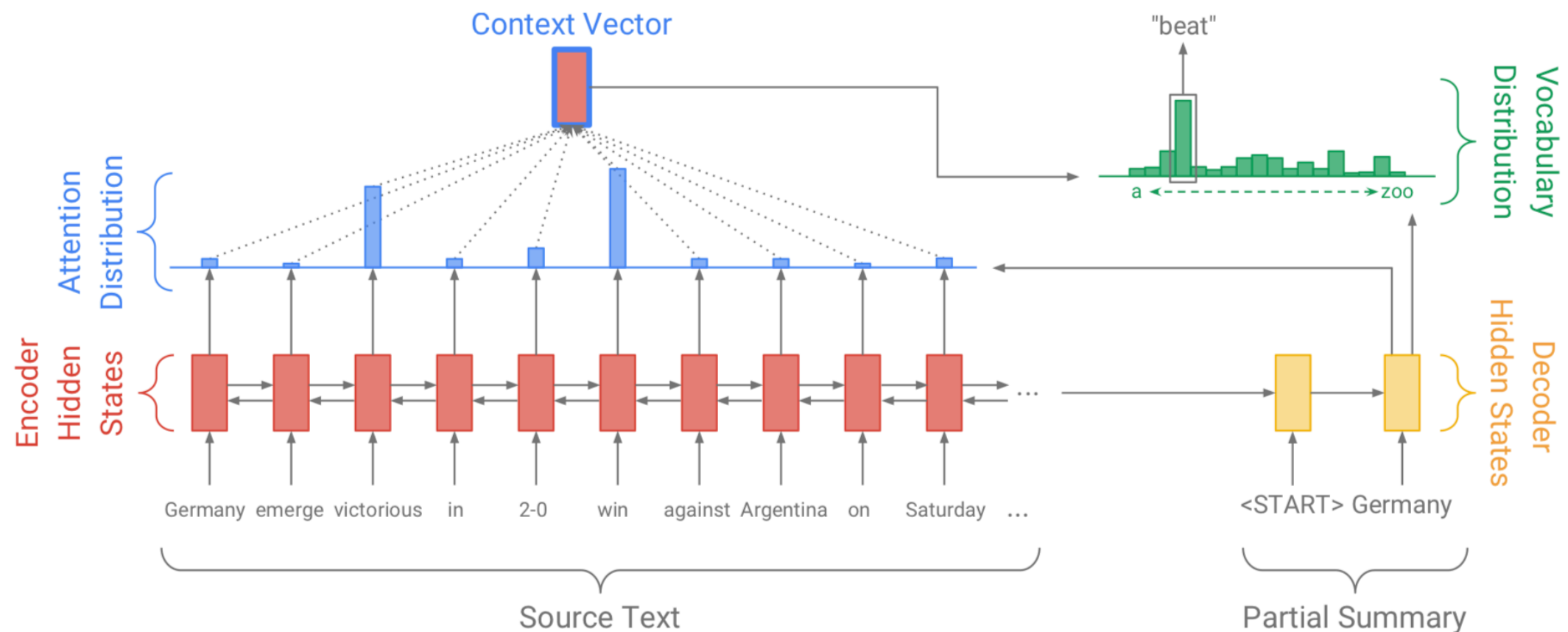


Figure 2: Baseline sequence-to-sequence model with attention. The model may attend to relevant words in the source text to generate novel words, e.g., to produce the novel word *beat* in the abstractive summary *Germany beat Argentina 2-0* the model may attend to the words *victorious* and *win* in the source text.

See et al. (2017) "Get To The Point: Summarization with Pointer-Generator Networks". ACL.

**... without a summary let's go
straight here ...**

... Outlook on InfoLing II

- This module covered some of the foundations of natural language processing
- There is *a lot* more to it (see, for example, the chapters we did not cover in the textbook)
- InfoLing II is the module to take to learn about this
- You should now have solid knowledge to do well in InfoLing II
- Same underlying theme(s) as we have seen so far
- Most exciting development in recent (< 5) years are neural approaches
- Looking forward to seeing you next term!

Reading

- Check out the relevant EACL+ECIR 2021 papers
- Feel free to look ahead in the Jurafsky & Martin textbook at the chapters corresponding to the application areas covered today

... and finally ...

... there is so much more to explore

DISCOVER

THE SCIENCES | MIND | TECHNOLOGY | HEALTH | ENVIRONMENT | PLANET EARTH

THE SCIENCES

Fruit Fly Brain Hacked For Language Processing

A simulated fruit fly brain has learnt to perform natural language processing tasks.

The Physics arXiv Blog | By The Physics arXiv Blog | January 28, 2021 10:32 PM



arXiv.org > cs > arXiv:2101.06887

Search...

Help | Advanced S

Computer Science > Computation and Language

[Submitted on 18 Jan 2021]

Can a Fruit Fly Learn Word Embeddings?

Yuchen Liang, Chaitanya K. Ryali, Benjamin Hoover, Leopold Grinberg, Saket Navlakha, Mohammed J. Zaki, Dmitry Krotov

The mushroom body of the fruit fly brain is one of the best studied systems in neuroscience. At its core it consists of a population of Kenyon cells, which receive inputs from multiple sensory modalities. These cells are inhibited by the anterior paired lateral neuron, thus creating a sparse high-dimensional representation of the inputs. In this