Natural Language Processing

for Music

Sergio Oramas (Pandora)

Outline

Section I: Natural Language Processing

1. Introduction to NLP
2. Semantics
3. Text Representation
4. Deep Learning Architectures
5. Contextualized Embeddings
6. Multimodality
7. Conversational Systems

Outline

Section II: Use cases in Music

1. Automatic Knowledge Base Construction
2. Music Recommendation with Knowledge Graphs
3. Multimodal Music Recommendation
4. Multimodal Music Genre Classification
5. NLP for Music Knowledge Discovery

Introduction to NLP

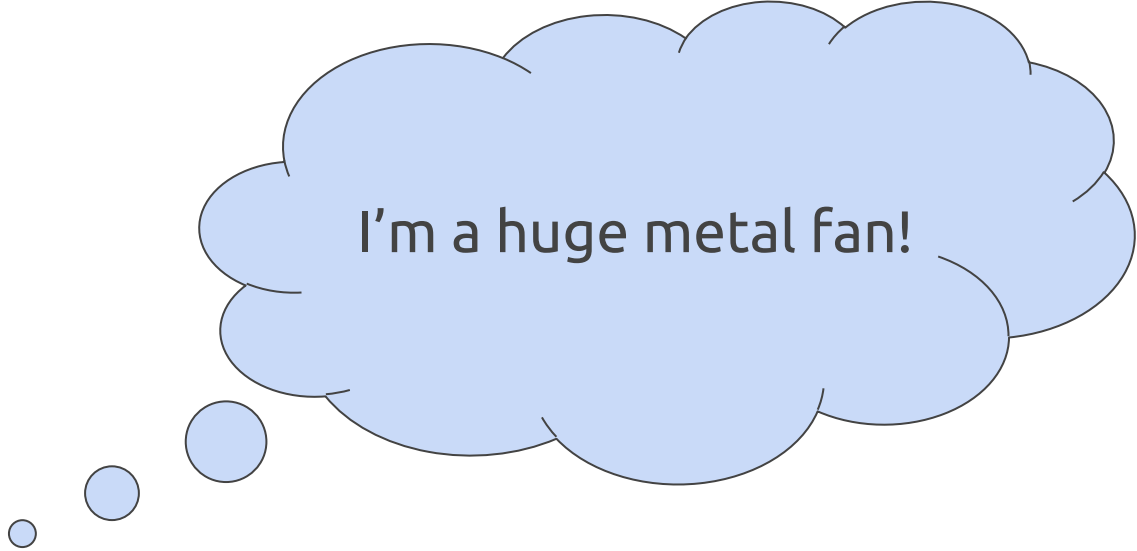
What is Natural Language Processing?

· NLP is a field of Computer Science and Artificial Intelligence concerned with the interaction between computers and human (natural) language.

· Alan Turing's paper *Computing Machinery and Intelligence* is believed to be the first NLP paper. It stated that a computer could be considered intelligent if it could carry on a conversation with a human being without the human realizing he/she were talking to a machine.

· NLP is a core component in daily life technologies: web search, speech recognition and synthesis, automatic summaries in the web, product (including music) recommendation, machine translation...

Why is it hard?



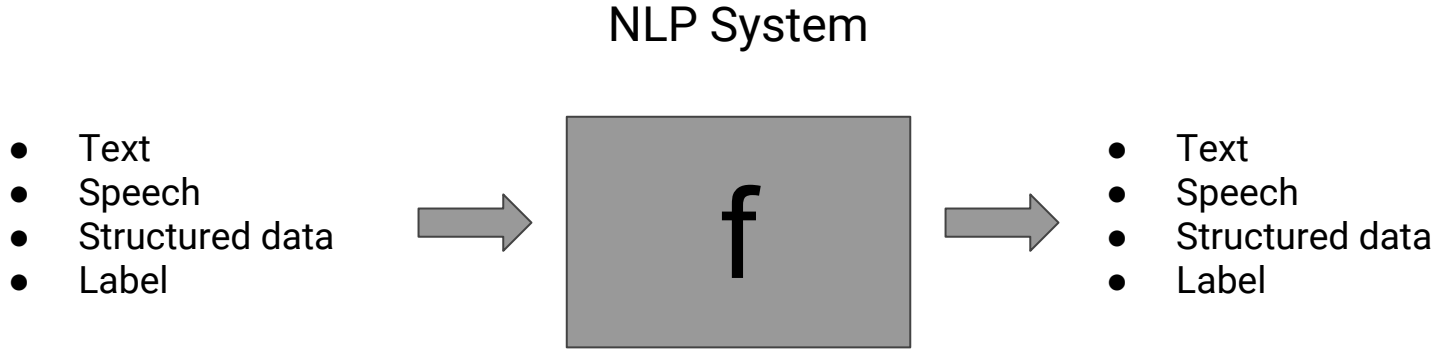
Why is it hard?

I’m a huge metal fan!

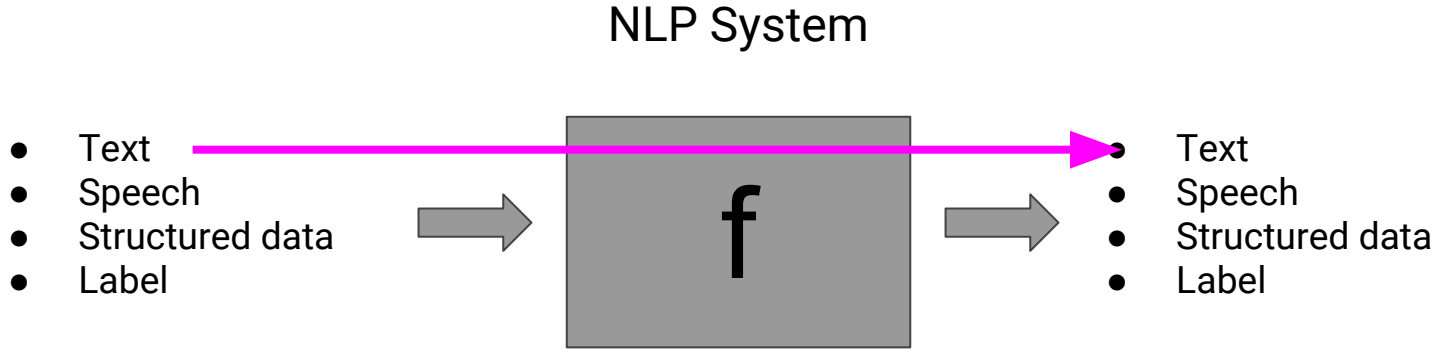
Why is it hard?

I’m a huge metal fan!

Natural Language Processing

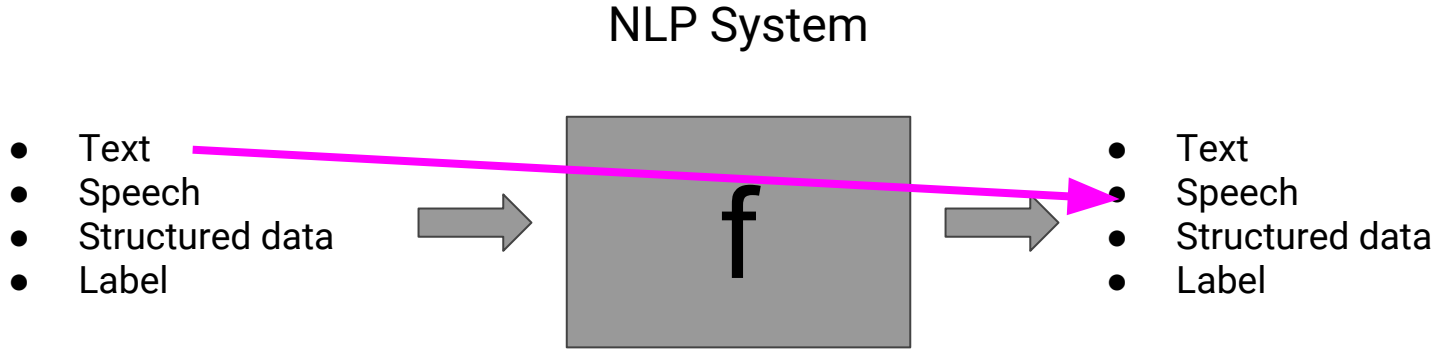


Natural Language Processing Tasks

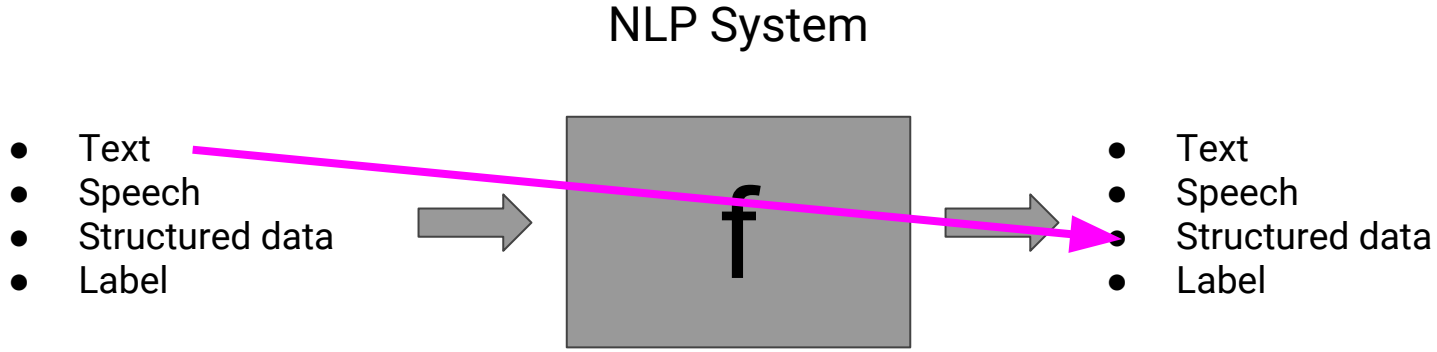


Summarization
Machine Translation
Dialogue
Question & Answering

Natural Language Processing Tasks

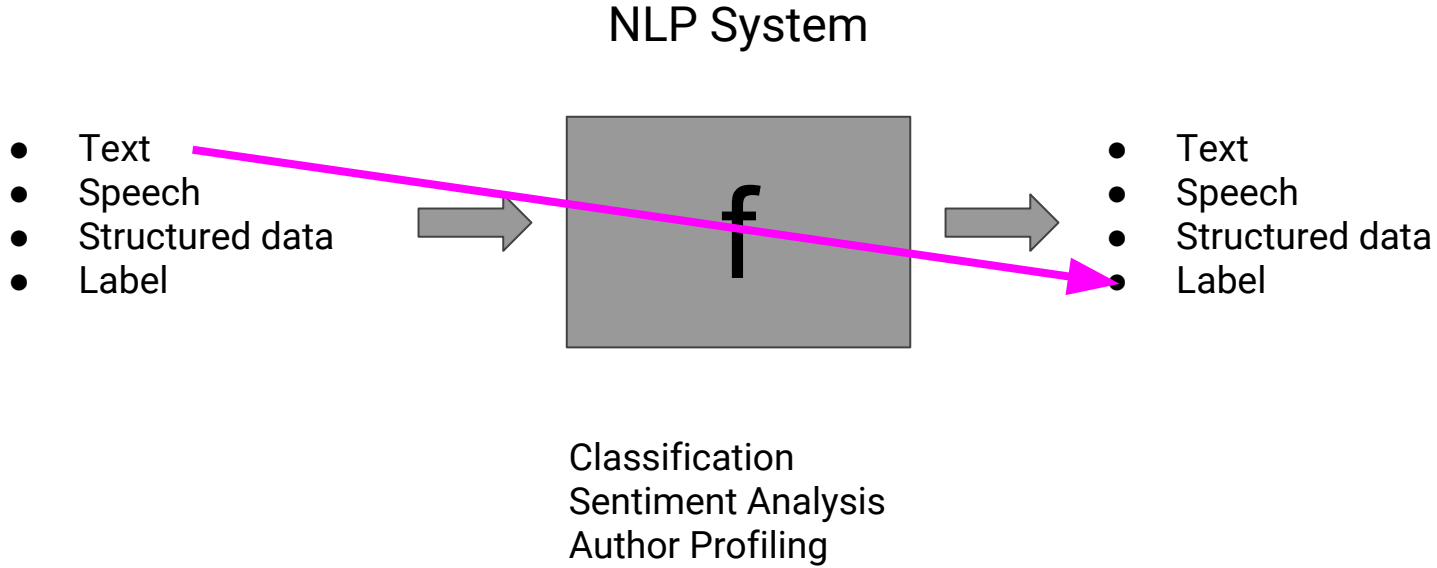


Natural Language Processing Tasks

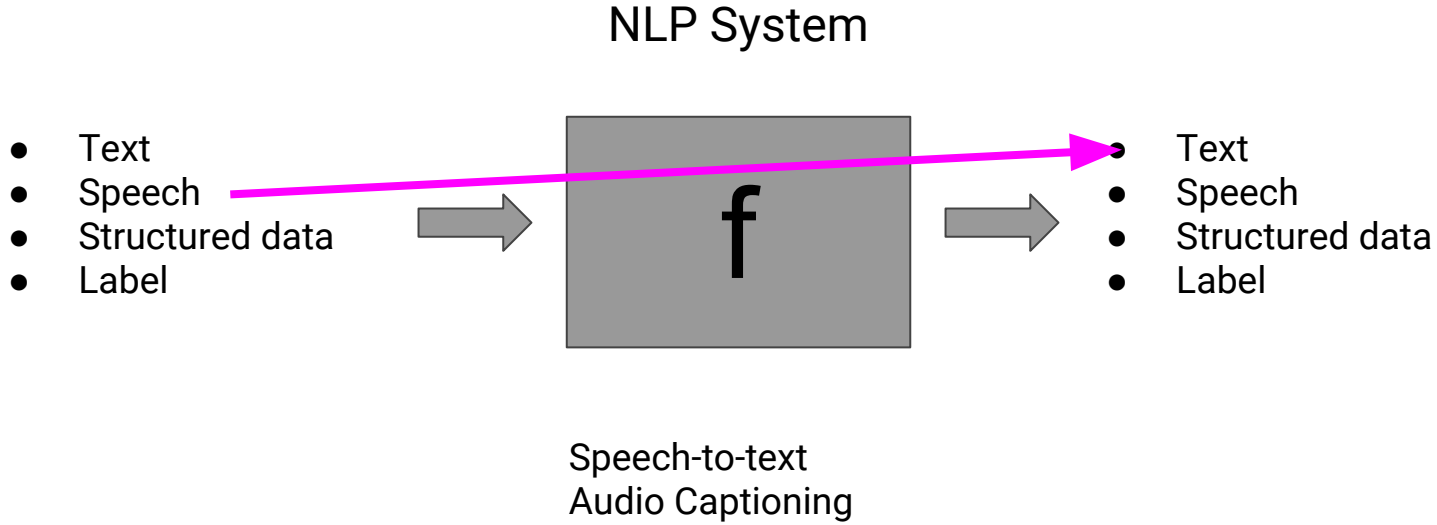


Relation Extraction
Entity Linking
Natural Language Understanding

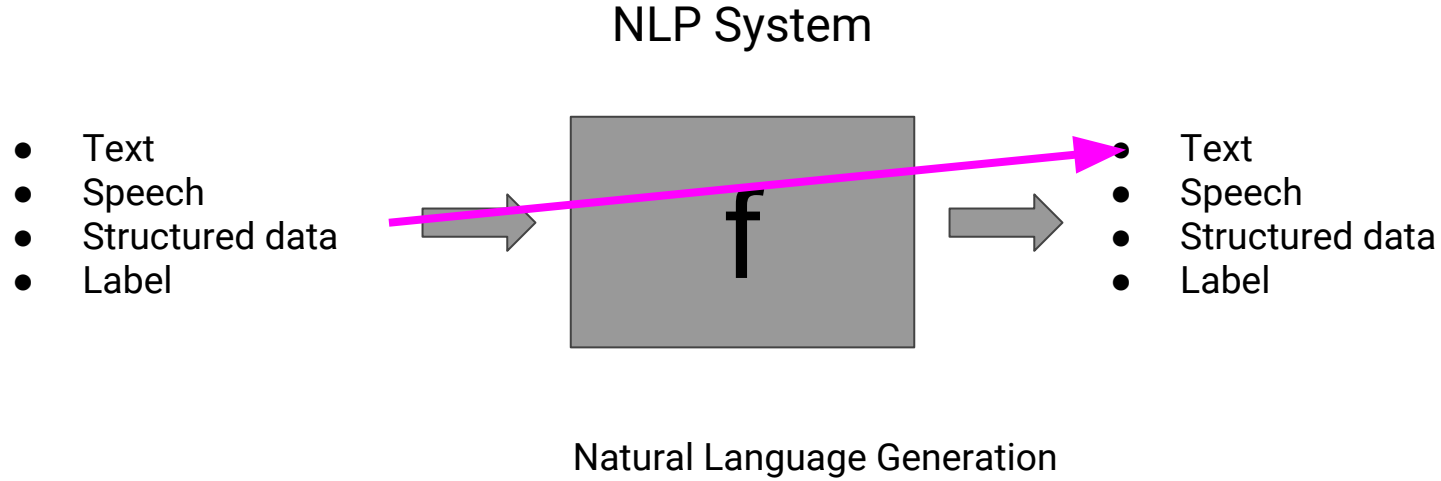
Natural Language Processing Tasks



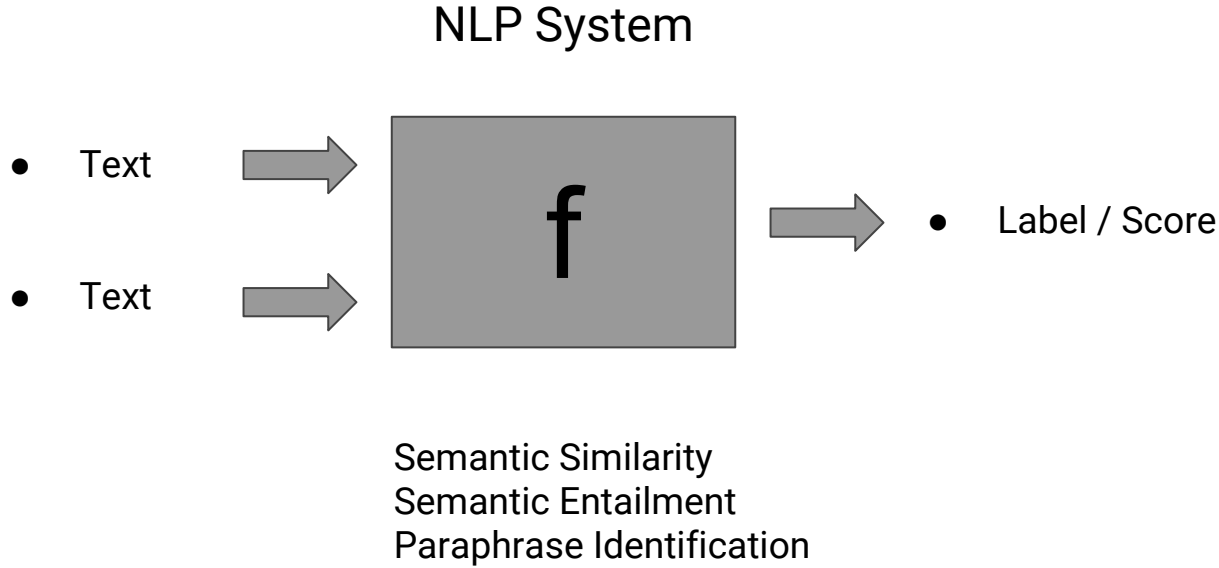
Natural Language Processing Tasks



Natural Language Processing Tasks



Natural Language Processing Tasks



Core Linguistic NLP Tasks

* Part-of-speech Tagging

* Syntactic Parsing

* Semantic Parsing

* Named Entity Recognition

* Coreference Resolution

* Word Sense Disambiguation

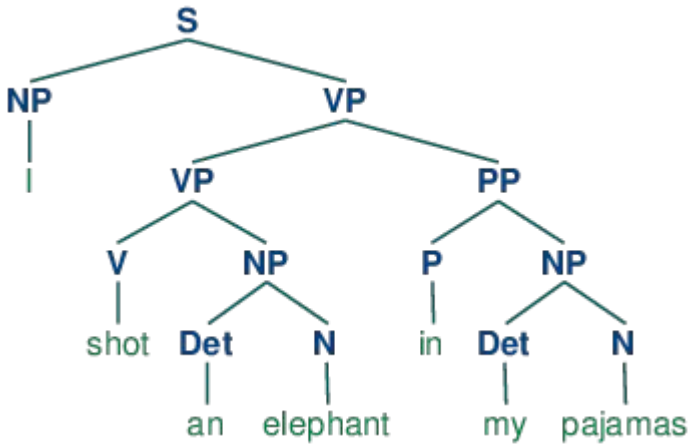
Core elements in NLP - **Part-of-Speech Tagging**

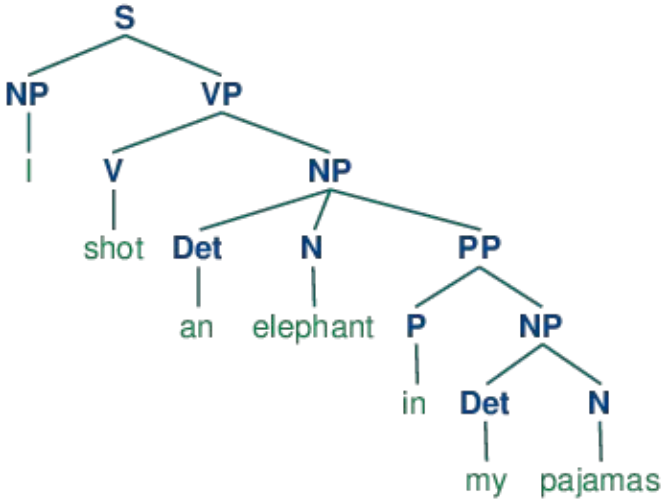
I like jazz music, it’s like being alive for a second.

Core elements in NLP - **Part-of-Speech Tagging**

Core elements in NLP - **Syntactic Parsing**

Identify relations holding between words or phrases in the sentence, and what is their *function*.

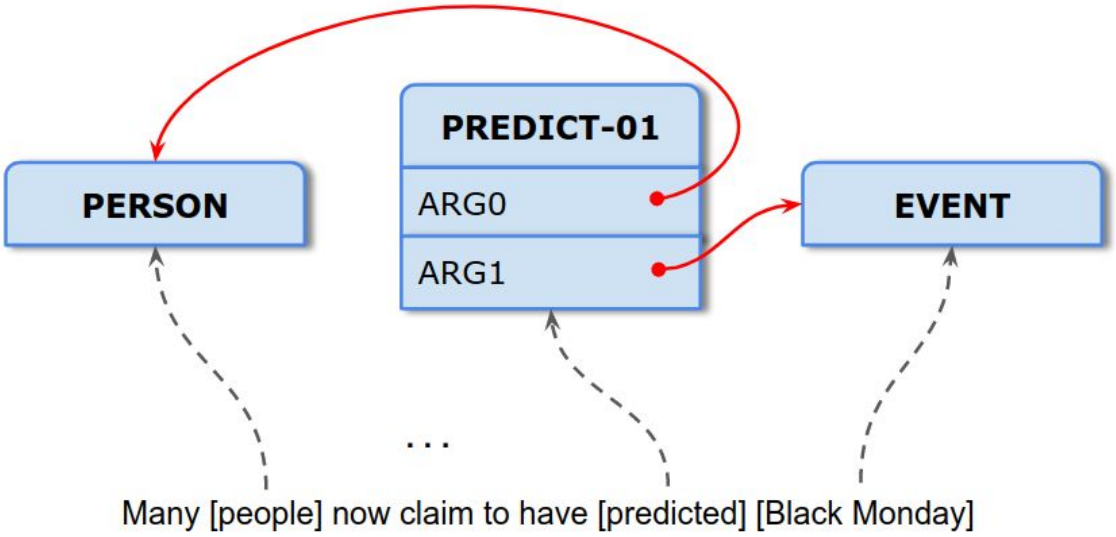




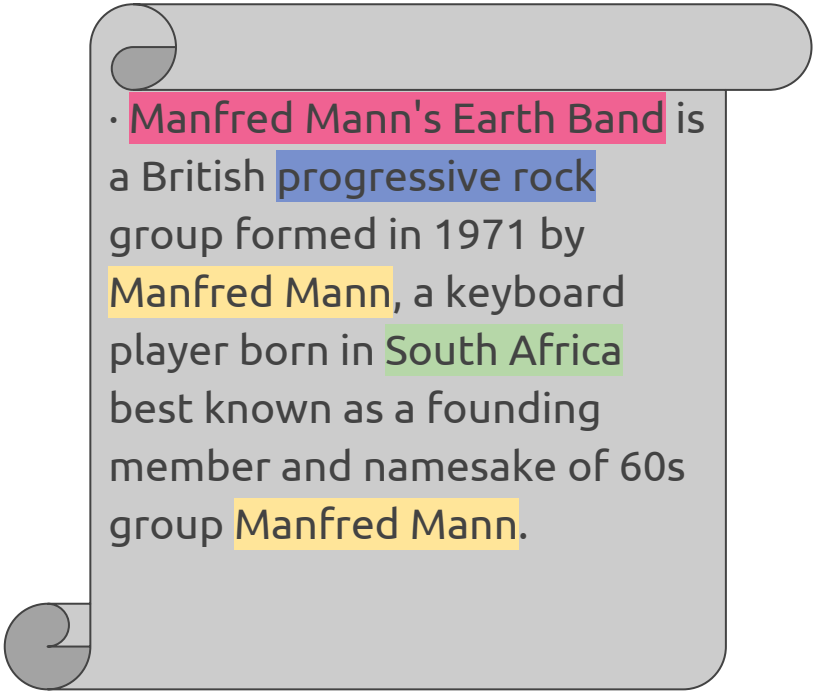
http://www.nltk.org/book/ch08.html

Core elements in NLP - **Semantic Parsing**

A level of parsing above morphology and syntax. Capture underlying semantics expressed in language. Most focus on verbs and their *arguments*.

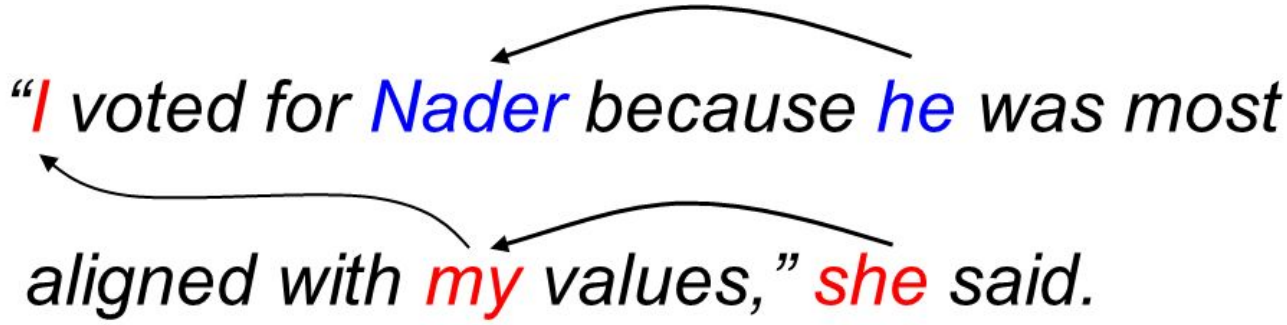


Core elements in NLP - **Named Entity Recognition**



Band
Music Genre
Artist
Country

Core elements in NLP - **Coreference Resolution**



http://nlp.stanford.edu/projects/coref.shtml

Core elements in NLP - **Word Sense Disambiguation**

· “The performance of that bass player was outstanding”

Core elements in NLP - **Word Sense Disambiguation**

· "The performance of that bass player was outstanding"

https://tackyraccoons.com/2011/11/21/all-your-bass-are-belong-to-us/

Tools

Alchemy API
http://www.alchemyapi.com/products/alchemylanguage/entity-extraction

AYLIEN API http://aylien.com/text-api

Stanford NLP http://nlp.stanford.edu:8080/sentiment/rntnDemo.html

Gensim python library https://radimrehurek.com/gensim/

Senti WordNet http://sentiwordnet.isti.cnr.it/

Software

Standalone

· OpenNLP: https://opennlp.apache.org/

· Stanford CoreNLP: http://stanfordnlp.github.io/CoreNLP/

· Freeling: http://nlp.lsi.upc.edu/freeling/node/1

· Gate: https://gate.ac.uk/

· Mate Parser: http://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/matetools.en.html

Python Libraries

· Spacy: https://spacy.io

· Pattern: http://www.clips.ua.ac.be/pattern

· NLTK: http://www.nltk.org/

· Gensim: https://radimrehurek.com/gensim/

· Blob: http://textblob.readthedocs.io/en/dev/

· Rake: https://www.airpair.com/nlp/keyword-extraction-tutorial

Software

ML toolkits/libraries widely used in NLP

· CRF++: https://taku910.github.io/crfpp/

· Mallet: http://mallet.cs.umass.edu/

· Networkx: https://networkx.github.io

· Weka: http://www.cs.waikato.ac.nz/ml/weka/

· **Deep Learning:**

- **Keras https://keras.io/**
- **Tflearn http://tflearn.org/**
- **Tensorflow https://www.tensorflow.org/**
- **Theano http://deeplearning.net/software/theano/**
- **DyNet (formerly cnn) https://github.com/clab/dynet**

References - NLP

Part-of-Speech Tagging: Schmid, H. (1994, September). Probabilistic part-of-speech tagging using decision trees. In Proceedings of the international conference on new methods in language processing (Vol. 12, pp. 44-49).

Parsing: Chomsky, N. (2002). Syntactic structures. Walter de Gruyter. ; Nivre, J. (2003). An efficient algorithm for projective dependency parsing. In Proceedings of the 8th International Workshop on Parsing Technologies (IWPT).

Named Entity Recognition: Tjong Kim Sang, E. F., & De Meulder, F. (2003, May). Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4 (pp. 142-147). Association for Computational Linguistics.

Coreference Resolution: Soon, W. M., Ng, H. T., & Lim, D. C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. Computational linguistics, 27(4), 521-544.

Summarization: Saggion, H., & Lapalme, G. (2002). Generating indicative-informative summaries with sumUM. Computational linguistics, 28(4), 497-526.

Simplification: Chandrasekar, R., Doran, C., & Srinivas, B. (1996, August). Motivations and methods for text simplification. In Proceedings of the 16th conference on Computational linguistics-Volume 2 (pp. 1041-1044). Association for Computational Linguistics.

Sentiment Analysis: Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

Author Profiling: Estival, D., Gaustad, T., Pham, S. B., Radford, W., & Hutchinson, B. (2007). Author profiling for English emails. In Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics (PACLING'07) (pp. 263-272).

Topic Modeling: Wallach, H. M. (2006, June). Topic modeling: beyond bag-of-words. In Proceedings of the 23rd international conference on Machine learning (pp. 977-984). ACM.

Machine Translation: Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., ... & Dyer, C. (2007, June). Moses: Open source toolkit for statistical machine translation. In Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions (pp. 177-180). Association for Computational Linguistics.

Lexical Semantics: Cruse, D. A. (1986). Lexical semantics. Cambridge University Press.

Word Sense Disambiguation. Navigli, R. (2009). "Word sense disambiguation: A survey." ACM Computing Surveys (CSUR) 41.2: 10.

References - KBs

WordNEt: Miller, George A. "WordNet: a lexical database for English." *Communications of the ACM* 38.11 (1995): 39-41.

Chebi: Degtyarenko, Kirill, et al. "ChEBI: a database and ontology for chemical entities of biological interest." Nucleic acids research 36.suppl 1 (2008): D344-D350.

Snomed: Spackman, Kent A., Keith E. Campbell, and Roger A. Côté. "SNOMED RT: a reference terminology for health care." Proceedings of the AMIA annual fall symposium. American Medical Informatics Association, 1997.

BabelNet: Navigli, Roberto, and Simone Paolo Ponzetto. "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network." Artificial Intelligence 193 (2012): 217-250.

DBPedia: Auer, Sören, et al. "Dbpedia: A nucleus for a web of open data." The semantic web. Springer Berlin Heidelberg, 2007. 722-735.

Yago: Suchanek, Fabian M., Gjergji Kasneci, and Gerhard Weikum. "Yago: a core of semantic knowledge." Proceedings of the 16th international conference on World Wide Web. ACM, 2007.

NELL: Carlson, Andrew, et al. "Toward an Architecture for Never-Ending Language Learning." AAAI. Vol. 5. 2010.

PATTY: Nakashole, Ndapandula, Gerhard Weikum, and Fabian Suchanek. "PATTY: a taxonomy of relational patterns with semantic types." *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, 2012.

WiseNet: Moro, Andrea, and Roberto Navigli. "WiSeNet: Building a Wikipedia-based semantic network with ontologized relations." Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, 2012.

DefIE: Delli Bovi, Claudio, Luca Telesca, and Roberto Navigli. "Large-Scale Information Extraction from Textual Definitions through Deep Syntactic and Semantic Analysis." Transactions of the Association for Computational Linguistics 3 (2015): 529-543.

KB-Unify: Bovi, Claudio Delli, Luis Espinosa Anke, and Roberto Navigli. "Knowledge Base Unification via Sense Embeddings and Disambiguation." Proceedings of EMNLP. 2015.

MusicBrainz: Swartz, Aaron. "Musicbrainz: A semantic web service." IEEE Intelligent Systems 17.1 (2002): 76-77.

Discogs: www.discogs.com

Grove Online: http://www.oxfordmusiconline.com/public/book/omo_gmo

References

Pang, B., & Lee, L. (2006). *Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 1(2). 91–231.*

Tata, S., & Di Eugenio, B. (2010). Generating Fine-Grained Reviews of Songs from Album Reviews. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, (July), 1376–1385.

Ruihai Dong, Michael P O'Mahony, and Barry Smyth (2014). Further Experiments in Opinionated Product Recommendation. In ICCBR'14, pages 110–124.

Oramas S., Espinosa-Anke L., Lawlor A., Serra X., Saggion H. (2016). *Exploring Music Reviews for Music Genre Classification and Evolutionary Studies*. 17th International Society for Music Information Retrieval Conference. ISMIR 2016.

Dominique Moisi. The Geopolitics of Emotion: How Cultures of Fear, Humiliation, and Hope are Reshaping the World. Anchor Books, New York, NY, USA, 2010.

Semantics

Knowledge Repositories and Knowledge Bases

· A Knowledge Base (KB) is a rich form of Knowledge Repository (KR), term coined to differentiate from traditional *databases*.

· The term Knowledge Base may be used to refer to lexical databases, ontologies, and any graph-like Knowledge Repository.

· KBs are essential for AI tasks such as reasoning, inference or semantic search. Also for Word Sense Disambiguation, Entity Linking, Semantics…

Knowledge Bases

· Hand-crafted KBs

→ From generic to domain-specific. E.g. **WordNet**, **CheBi**, **SnomedCT**.

· Integrative Projects

→ Unify in one single resource manually curated KRs and KBs.

⇒ **BabelNet** (originally, WordNet + Wikipedia), **DBPedia**, **Yago**…

· Open Information Extraction for KB construction

→ **NELL**, **PATTY**, **WiseNet**, **DefIE**, **KB-Unify**...

Music Knowledge Bases

- **MusicBrainz** and **Discogs**

  → Open encyclopedias of music metadata

  → MB is regularly published as Linked Data by the LinkedBrainz project.

- **Grove Music Online**

  → Music *scholar* encyclopedia

Information Extraction

Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents.

Unstructured vs. Structured

Information Extraction related Tasks

- Entity Linking
- Relation Extraction
- Knowledge Base Construction
- Knowledge Graphs

Information Extraction

Unstructured text

“Hate It Here” was written by Wilco frontman , Jeff Tweedy .

Information Extraction

Entity Identification

“Hate It Here” was written by Wilco frontman , Jeff Tweedy .

Information Extraction

Entity Recognition

Information Extraction

Wilco (disambiguation)

From Wikipedia, the free encyclopedia

Wilco is an American rock band.

Wilco may also refer to:

- Wilco (voice procedure), a radio procedure word, short for "**Wil**l **Co**mply"; origin of the term
- *Wilco (The Album)*, an album by the band Wilco, or the title song, "Wilco (The Song)"
- *Wilco: Learning How to Die*, a book about the band, by Greg Kot
- Wilco (farm supply cooperative), an American chain of agricultural cooperative stores
- Wilco (tree), *Anadenanthera colubrina*, a South American tree
- Wilkinson County, Georgia, sometimes abbreviated as "Wilco"
- Williamson County, Texas, sometimes abbreviated as "Wilco"
- WilcoHess, the chain of gas stations

Information Extraction

Entity Linking or Disambiguation

https://en.wikipedia.org/wiki/Wilco

Organization

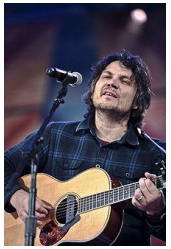
"Hate It Here" was written by Wilco frontman , Jeff Tweedy .

Work of art

http://musicbrainz.org/recording/246500ae-379b-4290-8716-d58b596753dd

Hate It Here
~ Recording by Wilco
MusicBrainz

Person

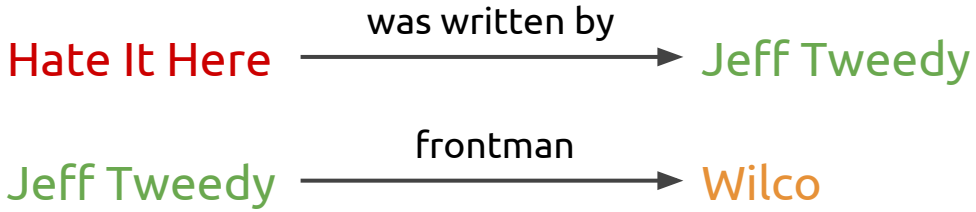


https://en.wikipedia.org/wiki/Jeff_Tweedy

Information Extraction

Relation Extraction

“Hate It Here” was written by Wilco frontman , Jeff Tweedy .



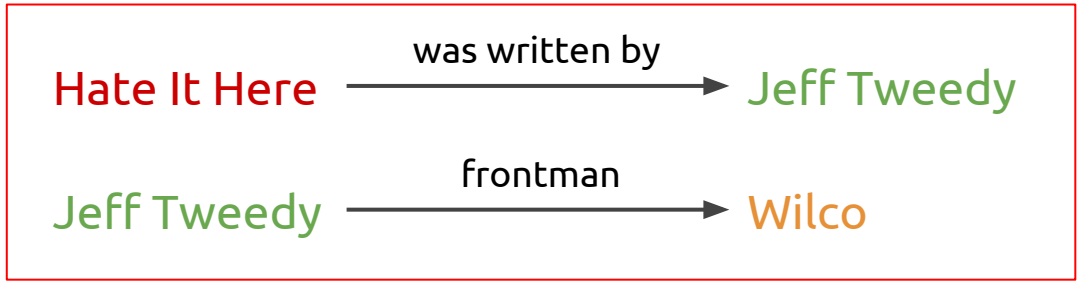
Information Extraction

Relation Extraction

Unstructured

"Hate It Here" was written by Wilco frontman , Jeff Tweedy .

Structured



Entity Linking

Entity linking is the task to associate, for a given candidate textual fragment, the most suitable entry in a reference Knowledge Base.

- Also referred to as Entity Disambiguation
- Typically Wikipedia, DBpedia, YAGO, Freebase as reference KB

Entity Linking

Entity linking needs to handle:

- Name variations (entities are referred to in many different ways)
  - e.g. Elvis, Elvis Presley, Elvis Aaron Presley, The King of Rock and Roll
- Entity ambiguity (the same string can refer to more than one entity)
  - e.g. Prince, Debut, Bach, Strauss
- Missing entities (there is no target entity in the knowledge base)
  - e.g. Supertrópica is not in Wikipedia

Entity Linking

Entity linking needs to handle:

- Name variations (entities are referre
  - e.g. Elvis, Elvis Presley, Elvis Aaron Presle
- Entity ambiguity (the same string ca
  - e.g. Prince, Debut, Bach, Strauss
- Missing entities (there is no target e
  - e.g. Supertrópica is not in Wikipedia

Entity Linking: Tools

Babelfy: Entity Linking + Word Sense Disambiguation. Web service. KB: BabelNet. http://babelfy.org/index

Tagme: Web service. KB: Wikipedia. https://tagme.d4science.org/tagme/

DBpedia Spotlight. Installable web service. KB: DBpedia. https://github.com/dbpedia-spotlight/dbpedia-spotlight

MEL: Music Entity Linking service. KB: MusicBrainz . http://mel.mtg.upf.edu

Relation Extraction

Detection and classification of semantic relations within a set of entities.

Numerous **variants**:

- Supervision: {fully, un, semi, distant}-supervision
- Undefined vs. pre-determined set of relations
- Binary vs. n-ary relations

Relation Extraction

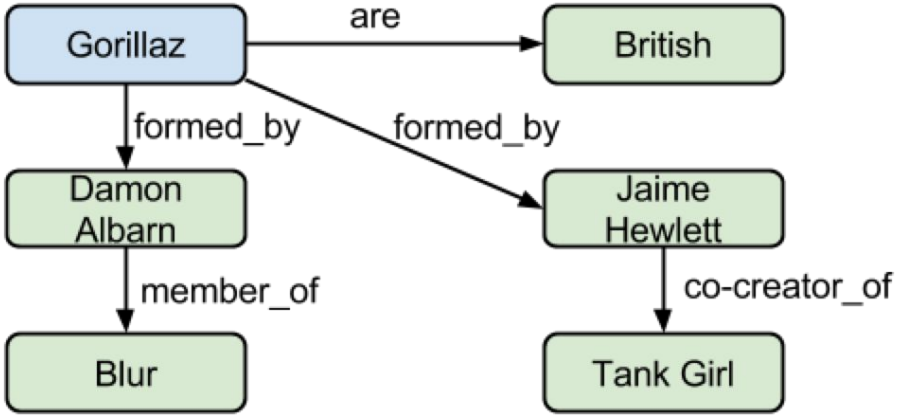
“Gorillaz are a British virtual band formed in 1998 by Damon Albarn of Blur, and Jaime Hawlett, co-creator of the comic book Tank Girl.”

Relation Extraction

“Gorillaz are a British virtual band formed in 1998 by Damon Albarn of Blur, and Jaime Hawlett, co-creator of the comic book Tank Girl.”

Relation Extraction

"Gorillaz are a British virtual band formed in 1998 by Damon Albarn of Blur, and Jaime Hawlett, co-creator of the comic book Tank Girl."



Relation Extraction (Tools)

ReVerb: OpenIE. Downloadable JAR. http://reverb.cs.washington.edu/

OpenIE: Successor of ReVerb. Downloadable JAR. http://openie.allenai.org/

DeepDive: Distant supervision. Installable python app.
http://deepdive.stanford.edu/

Relation Extraction (References)

Traditional IE

Zhao, S., & Grishman, R. (2005). Extracting relations with integrated information using kernel methods. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05* (pp. 419–426).

Weak Supervision

Bunescu, R. C., & Mooney, R. J. (2007). Learning to Extract Relations from the Web using Minimal Supervision. *Computational Linguistics*, *45*(June), 576–583.

Self Supervision

Carlson, A., Betteridge, J., & Kisiel, B. (2010). Toward an Architecture for Never-Ending Language Learning. *In Proceedings of the Conference on Artificial Intelligence (AAAI) (2010)*

Distant Supervision

Riedel, S., Yao, L., & McCallum, A. (2010). Modeling relations and their mentions without labeled text. In *Lecture Notes in Computer Science* (Vol. 6323 LNAI, pp. 148–163).

Relation Extraction (References)

Open IE

Fader, A., Soderland, S., & Etzioni, O. (2011). Identifying relations for open information extraction. *Proceedings of the Conference on Empirical Methods in Natural Language Processing EMNLP '11*, 1535–1545.

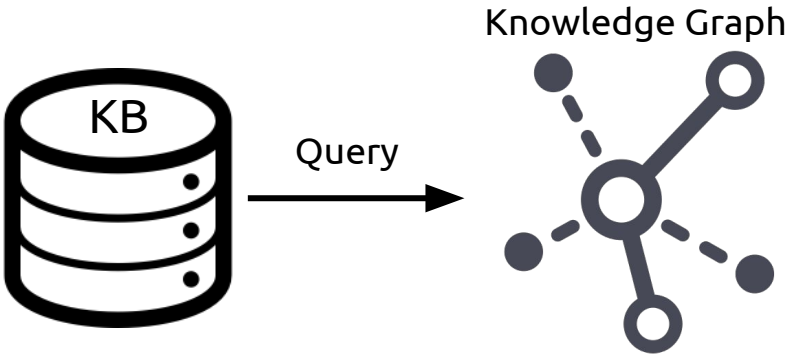
Semantic Open IE

Nakashole, N., Weikum, G., & Suchanek, F. M. (2012). PATTY: A Taxonomy of Relational Patterns with Semantic Types. *EMNLP-CoNLL*, (July), 1135–1145.

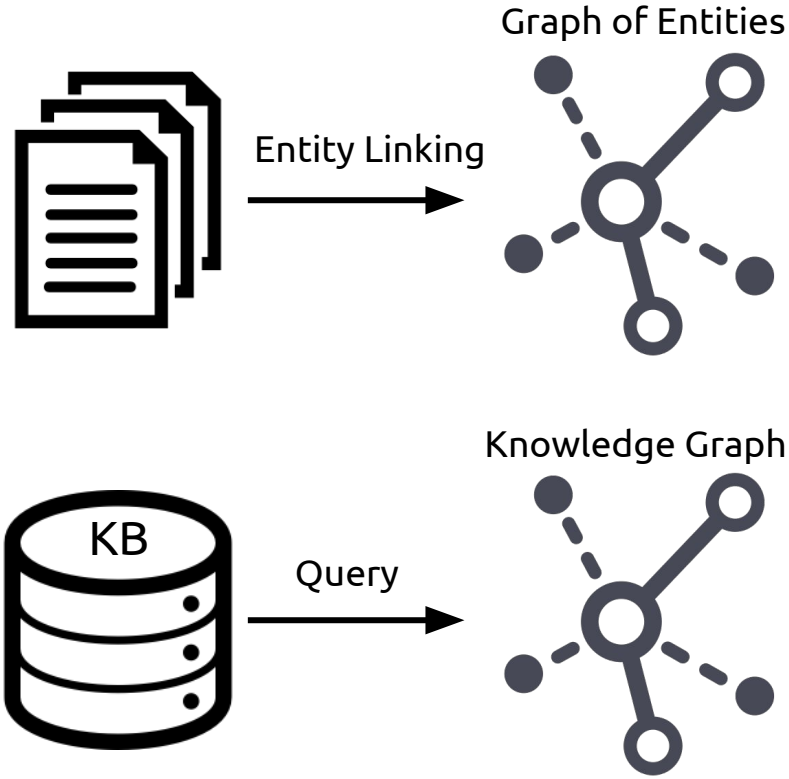
Delli Bovi, C., Telesca, L., & Navigli, R. (2015). Large-Scale Information Extraction from Textual Definitions through Deep Syntactic and Semantic Analysis. *Transactions of the Association for Computational Linguistics*, *3*, 529–543.

Oramas S., Espinosa-Anke L., Sordo M., Saggion H., Serra X. Information Extraction for Knowledge Base Construction in the Music Domain. Journal on Knowledge & Data Engineering, Elsevier.

Knowledge Graphs



Knowledge Graphs



Graph of Entities

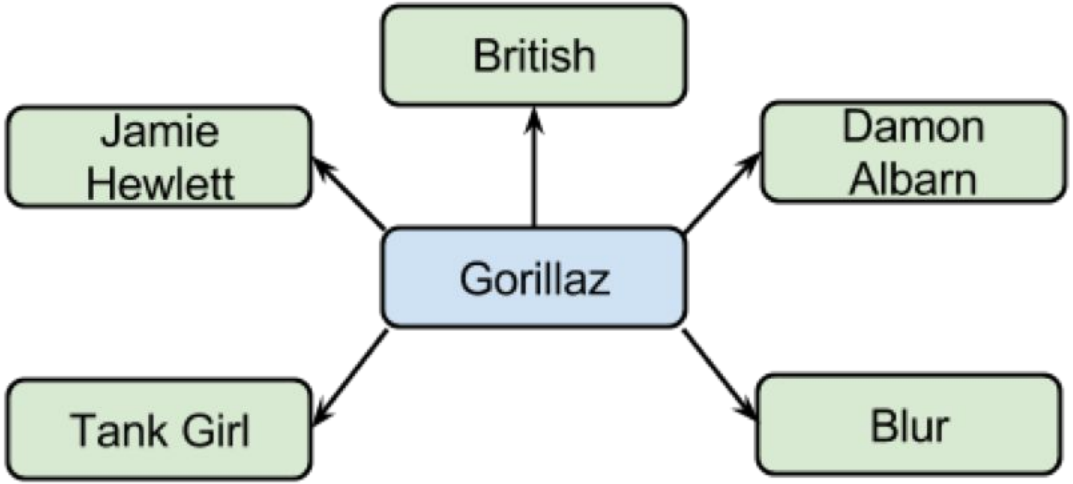
Gorillaz BIO: “Gorillaz are a British virtual band formed in 1998 by Damon Albarn of Blur, and Jaime Hawlett, co-creator of the comic book Tank Girl.”

Graph of Entities

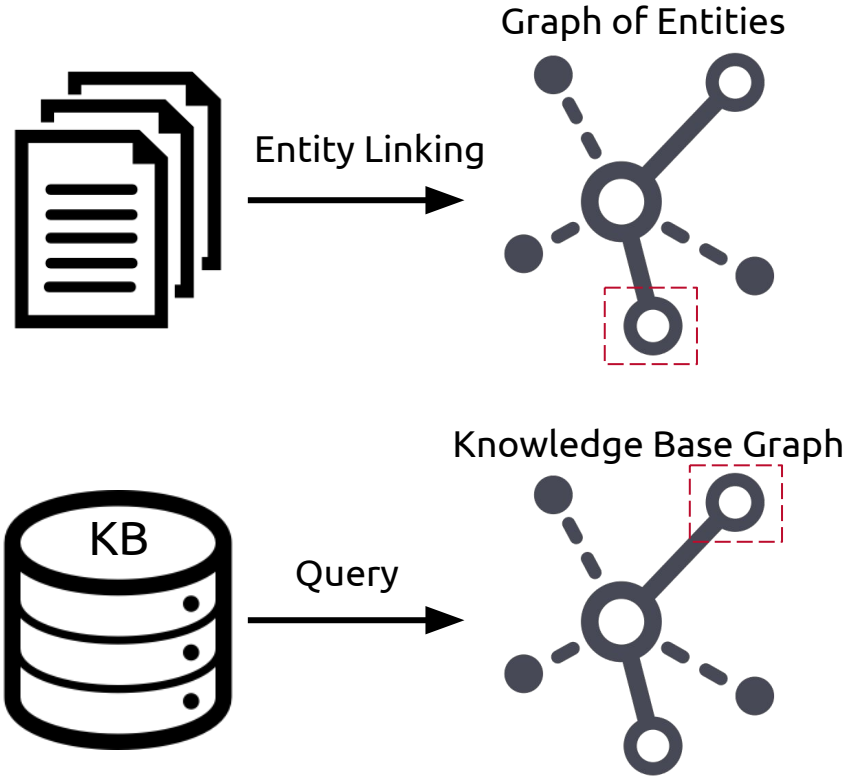
Gorillaz BIO: “Gorillaz are a British virtual band formed in 1998 by Damon Albarn of Blur, and Jaime Hawlett, co-creator of the comic book Tank Girl.”

Graph of Entities

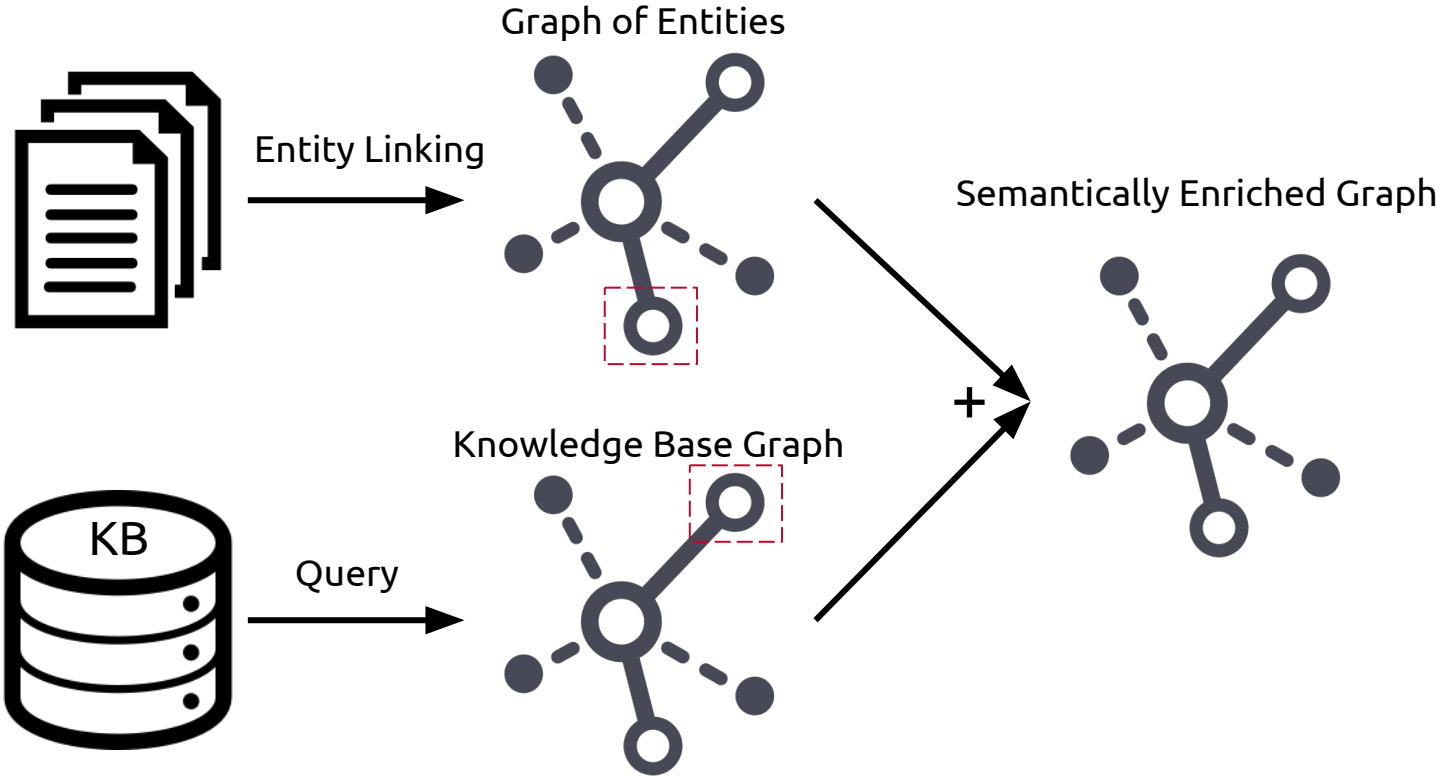
Gorillaz BIO: “Gorillaz are a British virtual band formed in 1998 by Damon Albarn of Blur, and Jaime Hawlett, co-creator of the comic book Tank Girl.”



Knowledge Graphs

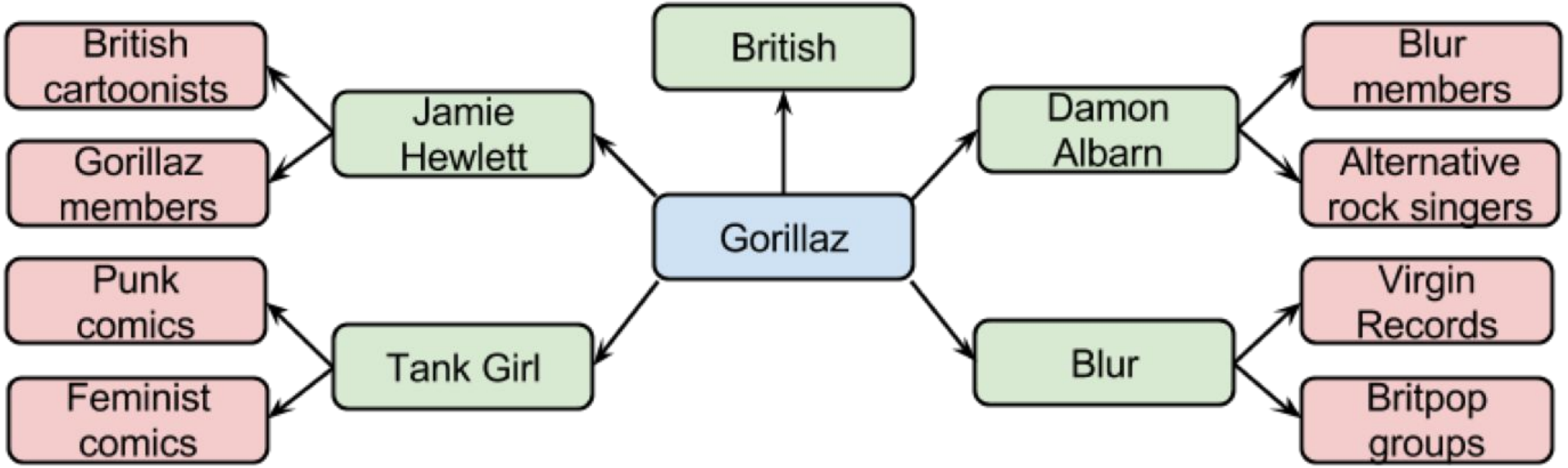


Knowledge Graphs



Semantically Enriched Graph

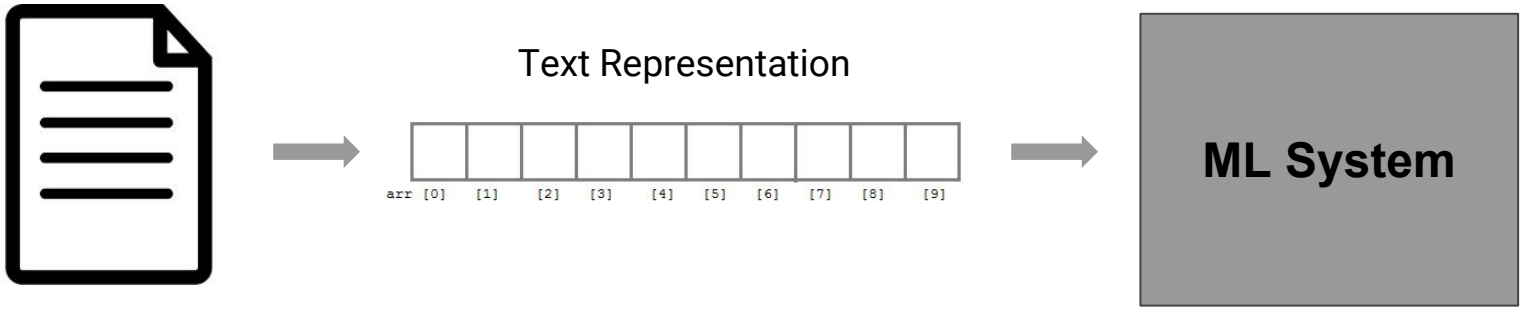
Gorillaz BIO: "Gorillaz are a British virtual band formed in 1998 by Damon Albarn of Blur, and Jaime Hawlett, co-creator of the comic book Tank Girl."



Text Representation

Text Representation

To process a text in a Machine Learning system we need to create a suitable data representation, which is typically a vector of real numbers.



Text Representation

Types of representation:

- Vector Space Model (VSM)
- Graph embeddings
- Word embeddings
- Char embeddings

Vector Space Model (VSM)

Bag-of-words (BoW): unordered list of words

Red Hot Chilli Peppers played at the red club

Vector Space Model (VSM)

VSM: Given the BoW extracted from the document, create a feature vector where each feature represents a word and the feature value is a term weight.

The term weight might be:

- a binary value (1 if word occurred in the document, and 0 if did not)
- a term frequency value (how many times the word occurred in the document)
- a numerical statistic value (e.g., tf-idf, BM25)

tf-idf

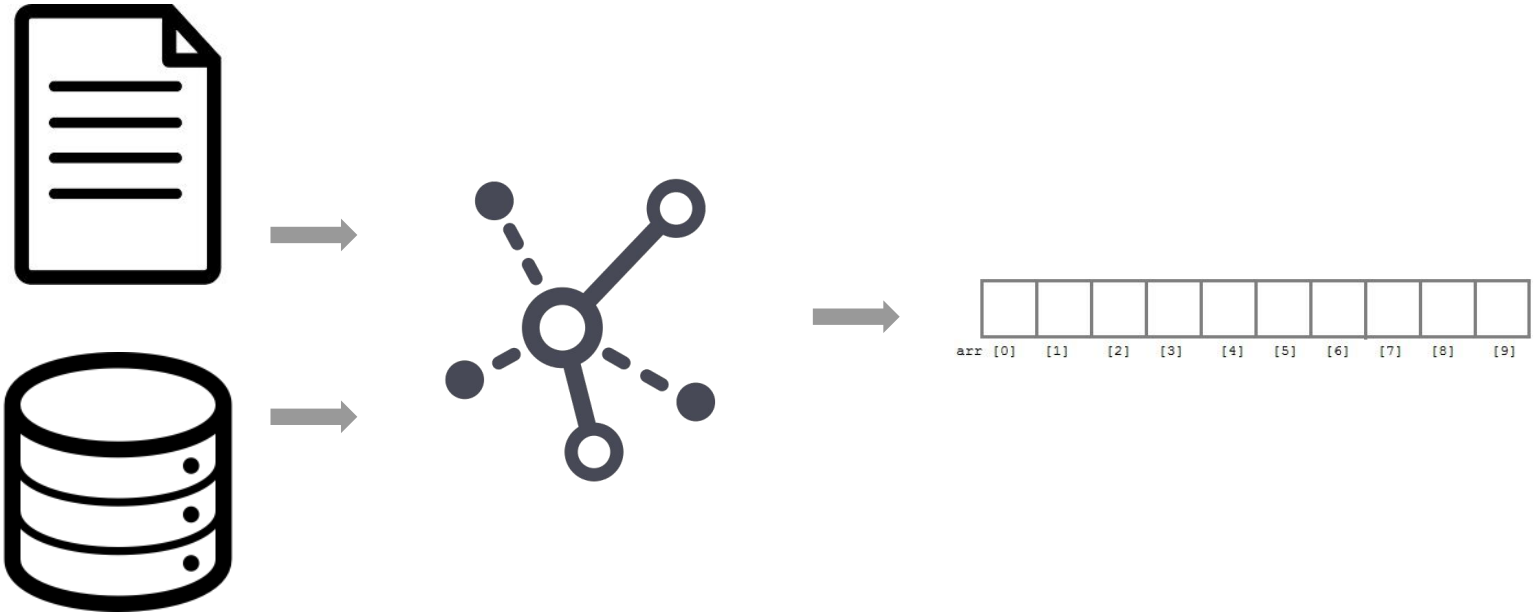
tf-idf: term frequency–inverse document frequency

- Numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus
- Widely used for Information Retrieval (search) and text-mining
- Frequently found in MIR literature

TF(t) = (Number of times term t appears in a document) / (Total number of terms in the document)

IDF(t) = log_e(Total number of documents / Number of documents with term t in it)

Graph Embedding



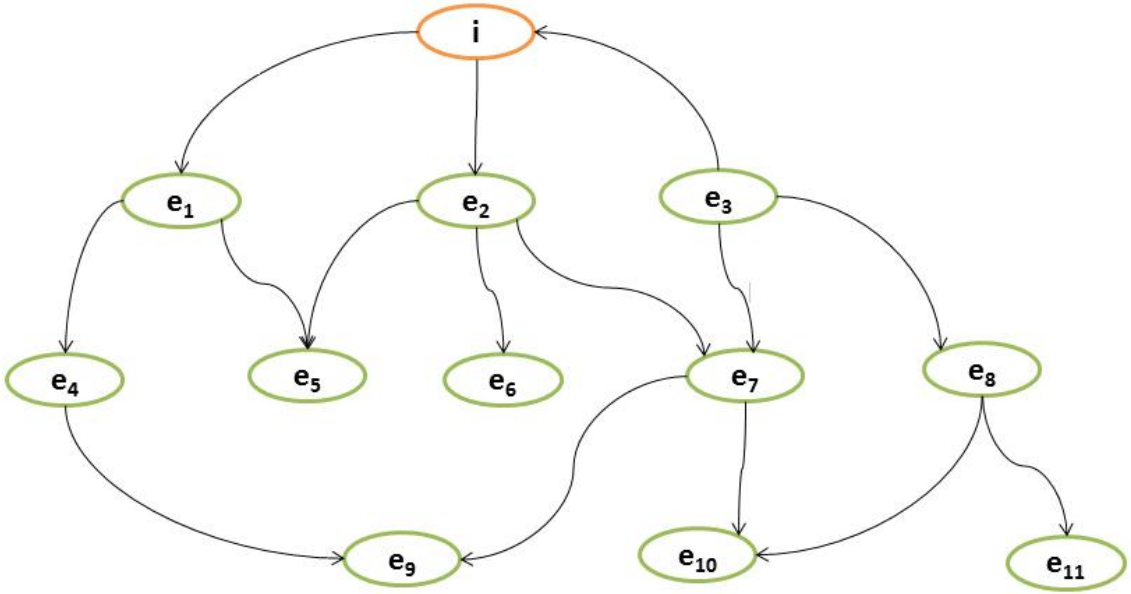
Graph Embeddings

Represent in a linear vector the structure of the graph

Multiple approaches:

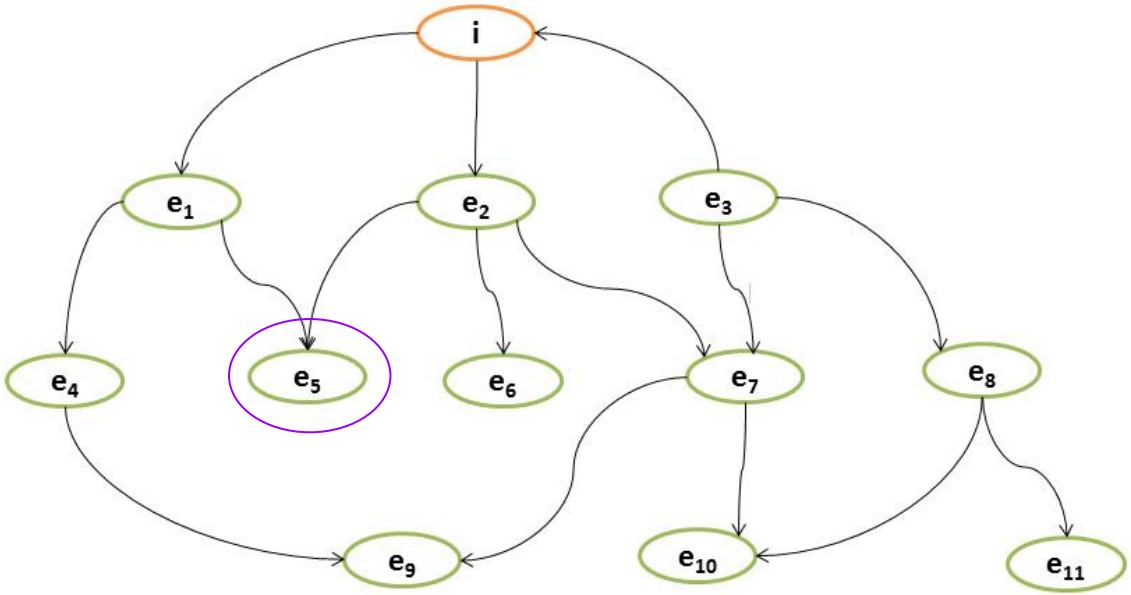
- Node and path embeddings
- Adjacency matrix
- Statistical models (e.g., node2vec)

Node-based Embedding



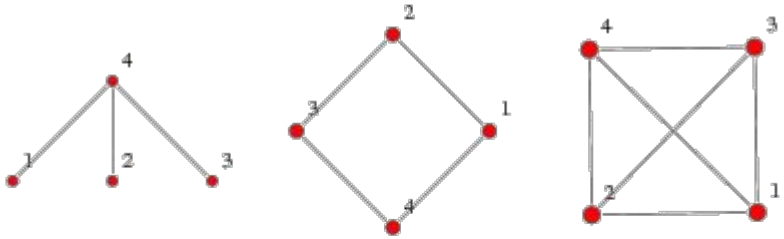
- One feature per entity
- Weight according to:
  - Distance to root
  - Number of in-links

Node-based Embedding



- One feature per entity
- Weight according to:
  - Distance to root
  - Number of in-links

Adjacency Matrix

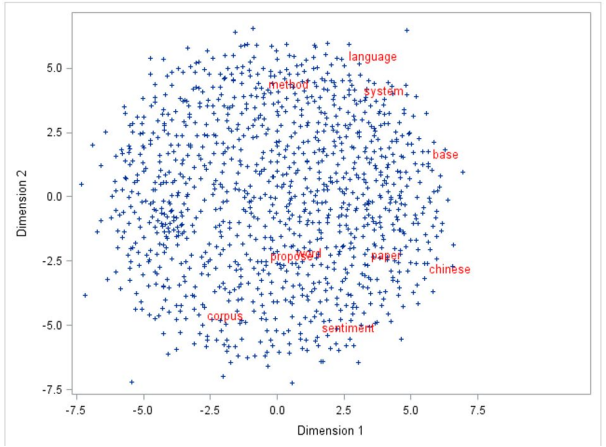


$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

Statistical models

Each node becomes a point in an N-dimensional space (node2vec)

Model



Word Embeddings

Distributional Semantics

“You shall know a word by the company it keeps”, Firth (1957).

Distributional Hypothesis: words that appear in similar contexts exhibit similar semantics.

He filled the **wampimuk** with the substance, passed it around we all drunk some.

We found a little, hairy **wampimuk** sleeping behind the tree.

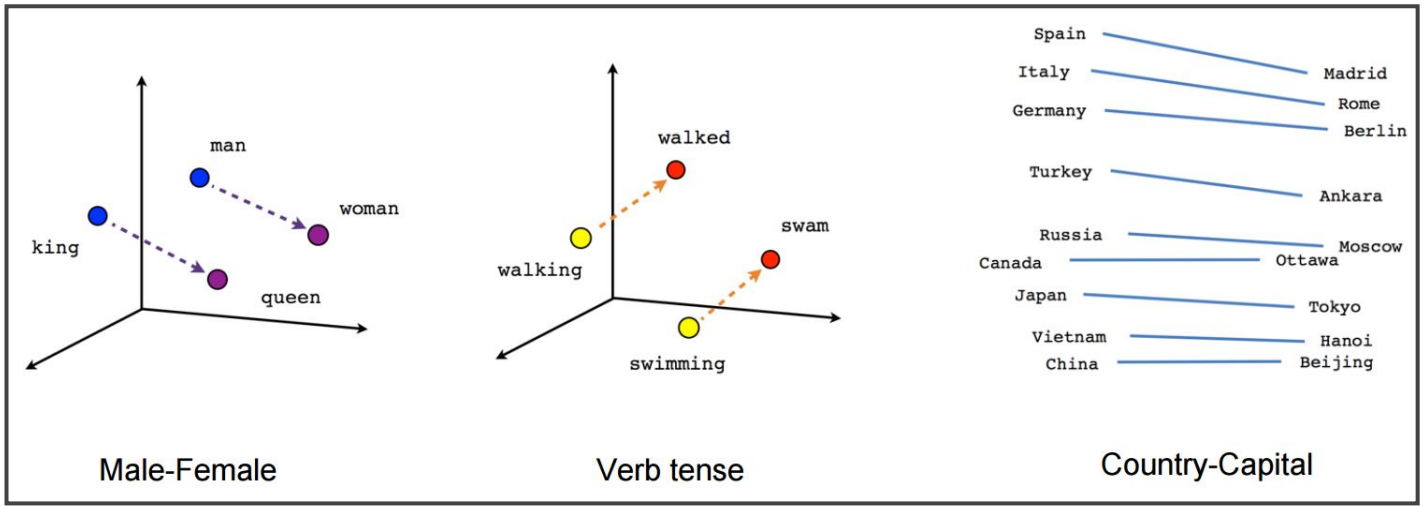
(McDonald and Ramscar, 2001)

Word Embeddings

Properties

Objective: Project linguistic items into a vector space

Similarity and geometrical properties



Word Embeddings

Word Embeddings

Pre-trained on external large corpora

Used as features in different tasks and domains

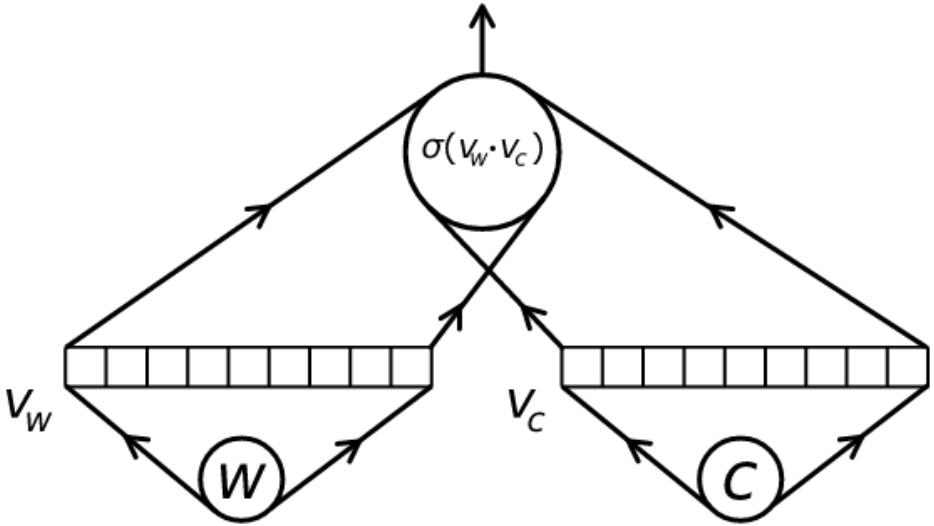
Popular approaches

- word2vec (Mikolov et al., 2013)
- Glove (Pennington et al., 2014)
- fastText (Bojanowski et al., 2017)

Word2vec

Predict a context word $\mathbf{c} \in (\mathbf{w}_{i-L}, \ldots, \mathbf{w}_{i-1}, \mathbf{w}_{i+1}, \ldots, \mathbf{w}_{i+L})$ given a word $\mathbf{w}_i$

$$P(D = 1|w, c) = \sigma(v_w \cdot v_c)$$



$$\sigma(a) = \frac{1}{1+e^{-a}}$$

More about Word2vec

https://www.tensorflow.org/versions/r0.10/tutorials/word2vec/index.html

http://cgi.cs.mcgill.ca/~enewel3/posts/implementing-word2vec/

http://hduongtrong.github.io/2015/11/20/word2vec/

T. Mikolov et al (2013): Distributed Representations of Words and Phrases and their Compositionality. Advances in neural information processing systems.

O. Levy, Y. Goldberg (2014): Neural Word Embedding as Implicit Matrix Factorization. NIPS 2014

Distributional Lexical Semantics

```
>>> from gensim.models import Word2Vec

>>> model = Word2Vec.load(PATH)
```

https://radimrehurek.com/gensim/models/word2vec.html

Word2vec in Playlists

Trained with Gensim in Art of the Mix playlists
(http://labrosa.ee.columbia.edu/projects/musicsim/aotm.htm)

model.most_similar('miles davis')
[('john clotrane', 0.88384414), ('dizzie gillespie', 0.78484219), ('charlie walker', 0.74520659)]

model.most_similar('marilyn manson')
[('godsmack', 0.93274206), ('white zombie', 0.91064525), ('drowning pool', 0.90275443)]

model.most_similar('nirvana')
[('soundgarden', 0.84231329), ('pearl jame', 0.8271907), ('oysterhead', 0.81855756)]

Character Embeddings

Each character has its own embedding

Word embeddings are combinations of character embeddings

Useful for out-of-vocabulary words

Deep Learning Architectures

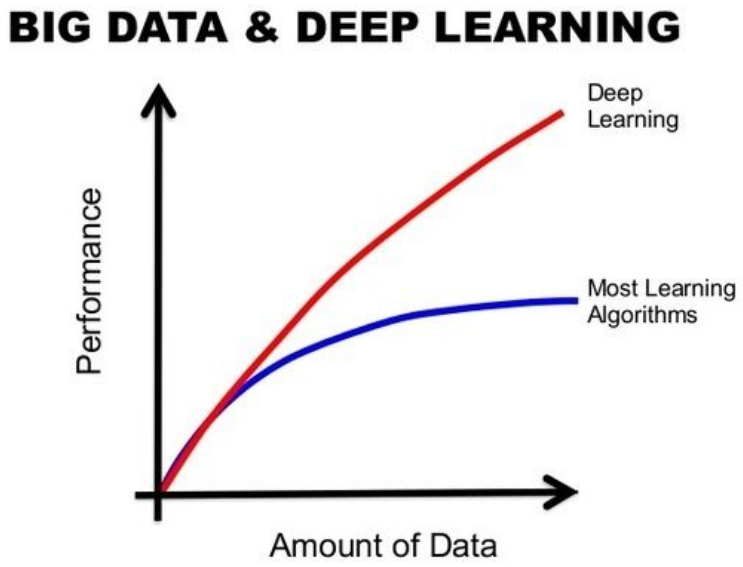
Deep Learning in Natural Language Processing

Deep Learning improves almost all tasks in NLP!! (as in many other fields)

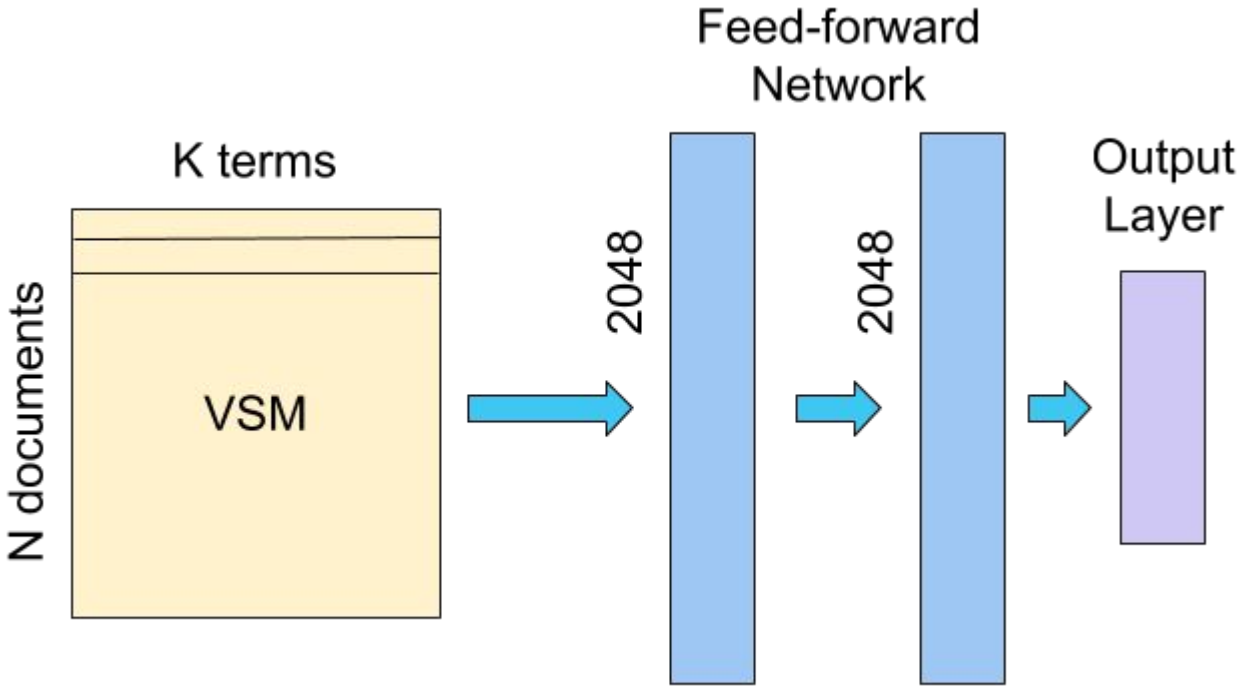
Deep Network Architectures: LSTM y CNN

More than words: character embeddings,

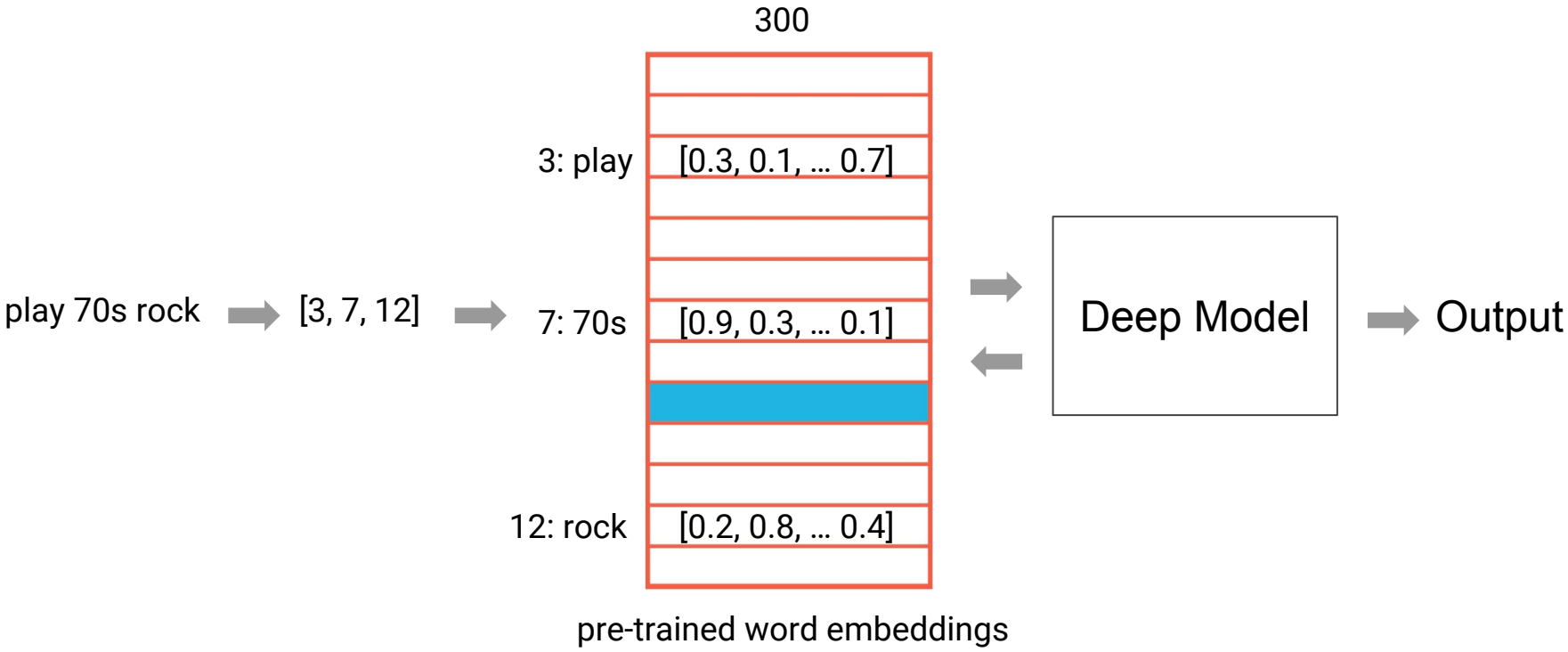
word embeddings, transfer learning



Feedforward Networks

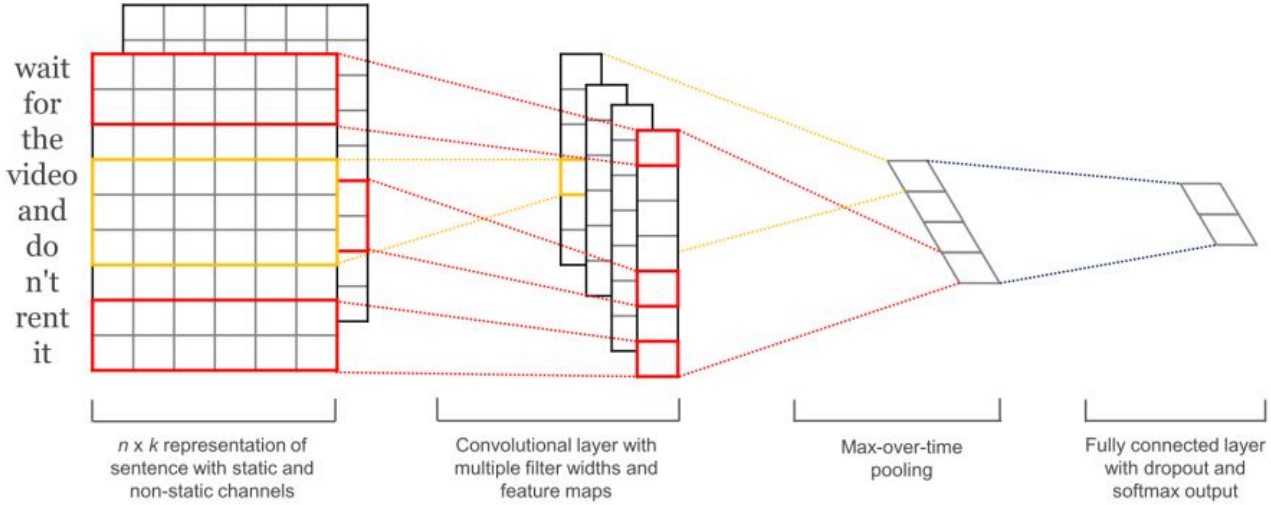


Embedding Layer

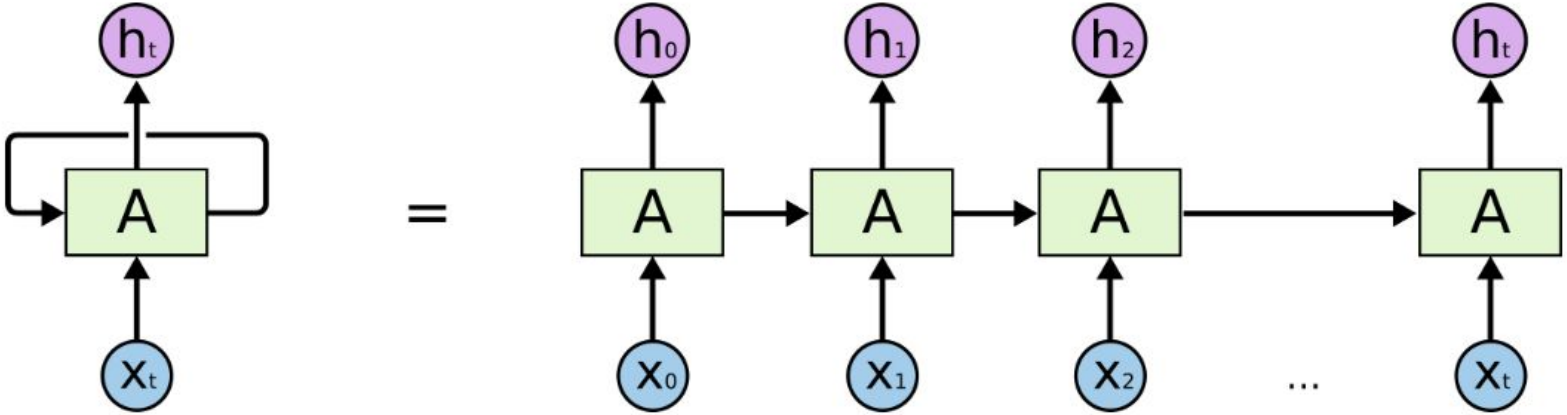


CNNs

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.



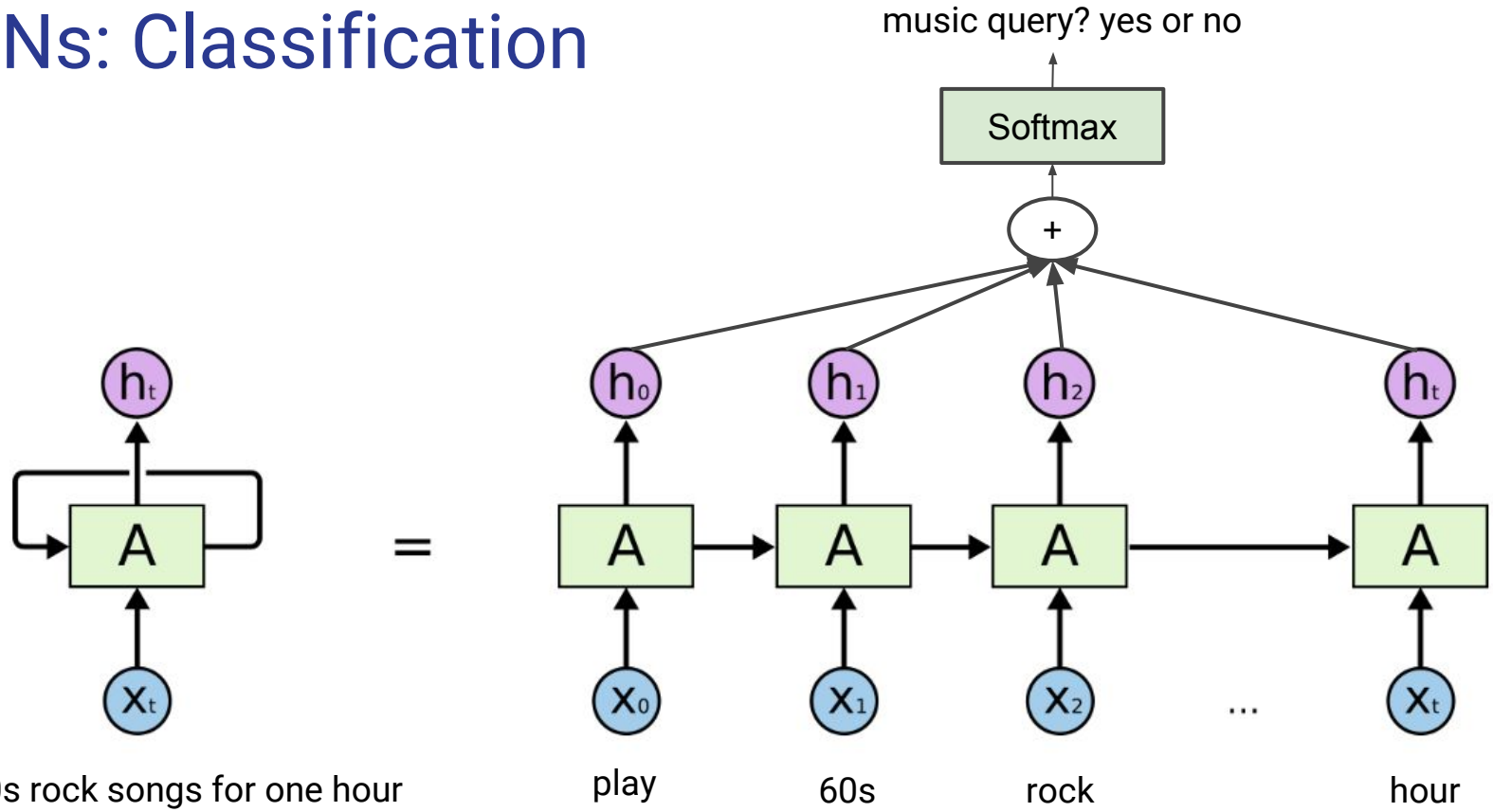
RNNs (LSTMs, GRUs, vanilla RNNs)



Play 60s rock songs for one hour

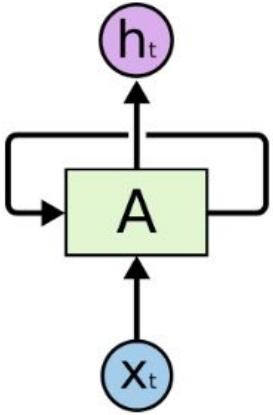
play 60s rock hour

RNNs: Classification

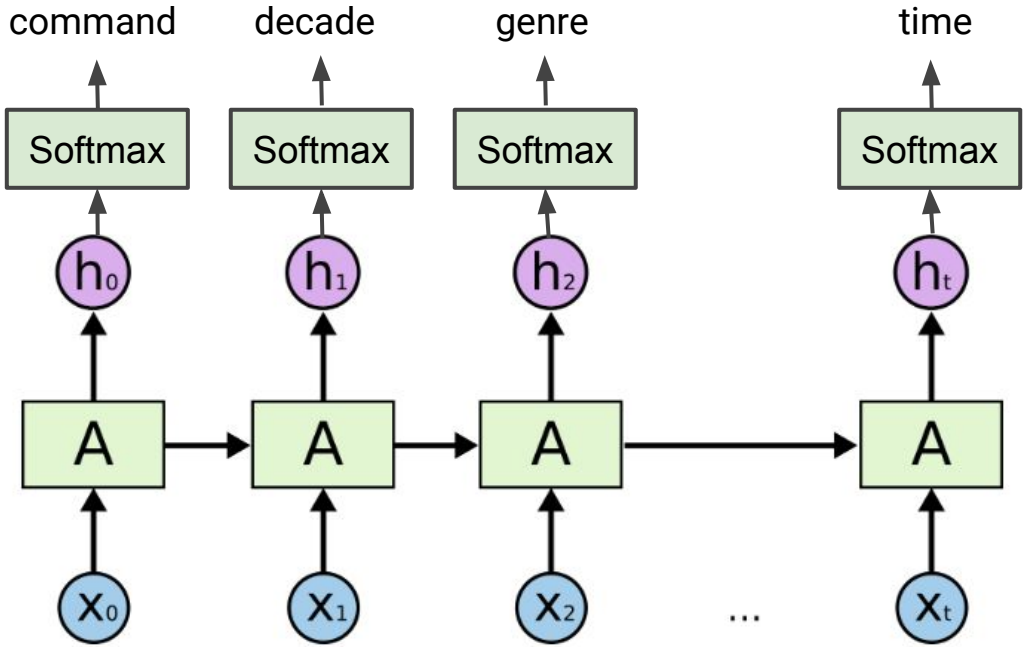


Play 60s rock songs for one hour

RNNs: Seq2seq

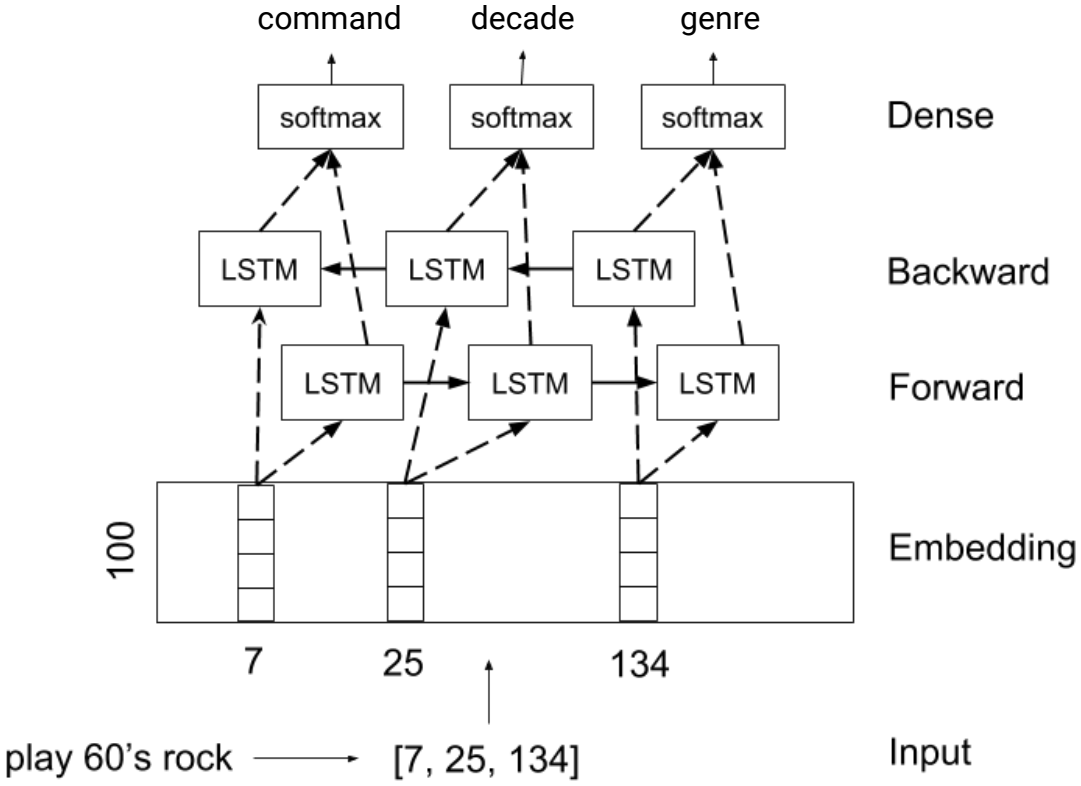


=

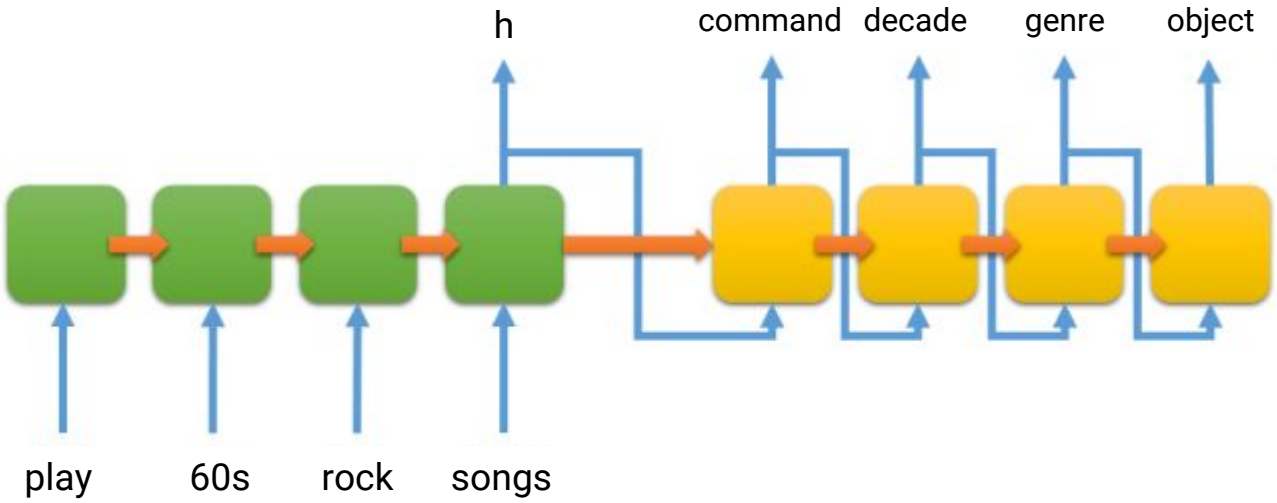


Play 60s rock songs for one hour

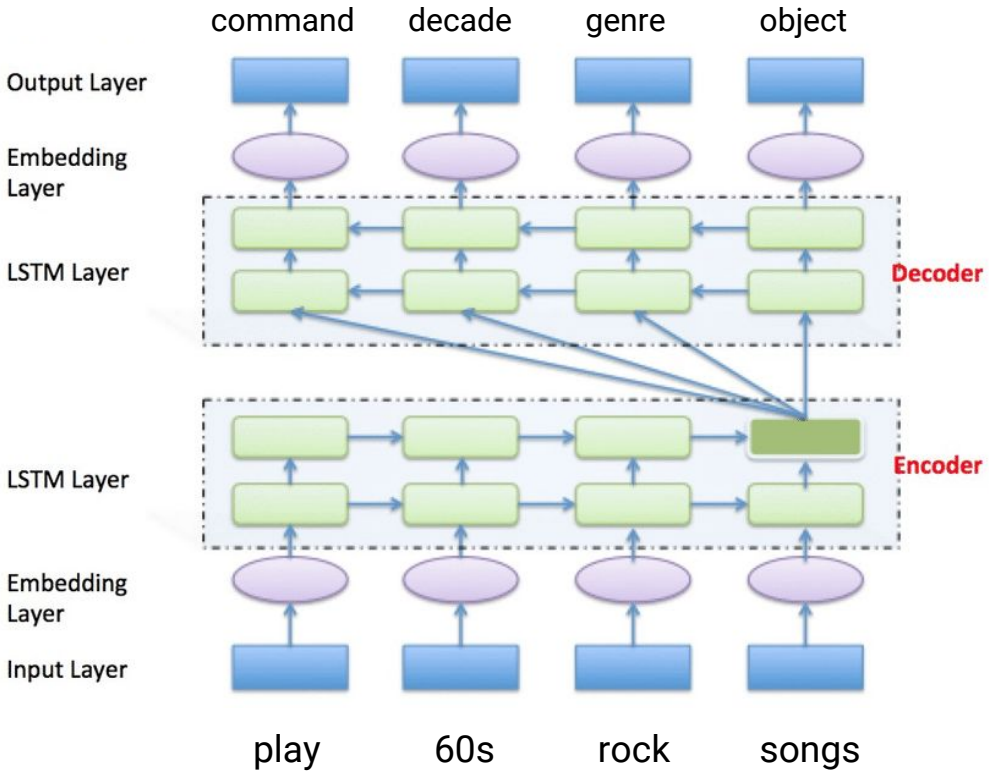
BiLSTM


command
decade
genre
softmax
softmax
softmax
Dense
LSTM
LSTM
LSTM
Backward
LSTM
LSTM
LSTM
Forward
100
Embedding
7
25
134
play 60's rock
[7, 25, 134]
Input

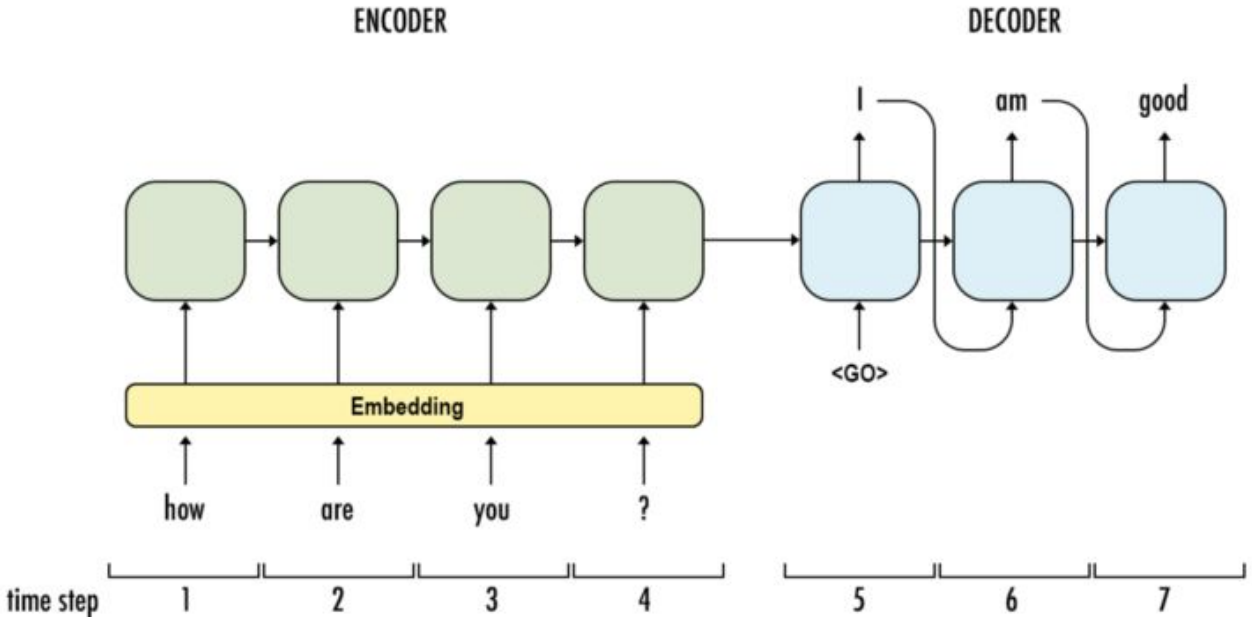
Encoder-Decoder



Encoder-Decoder

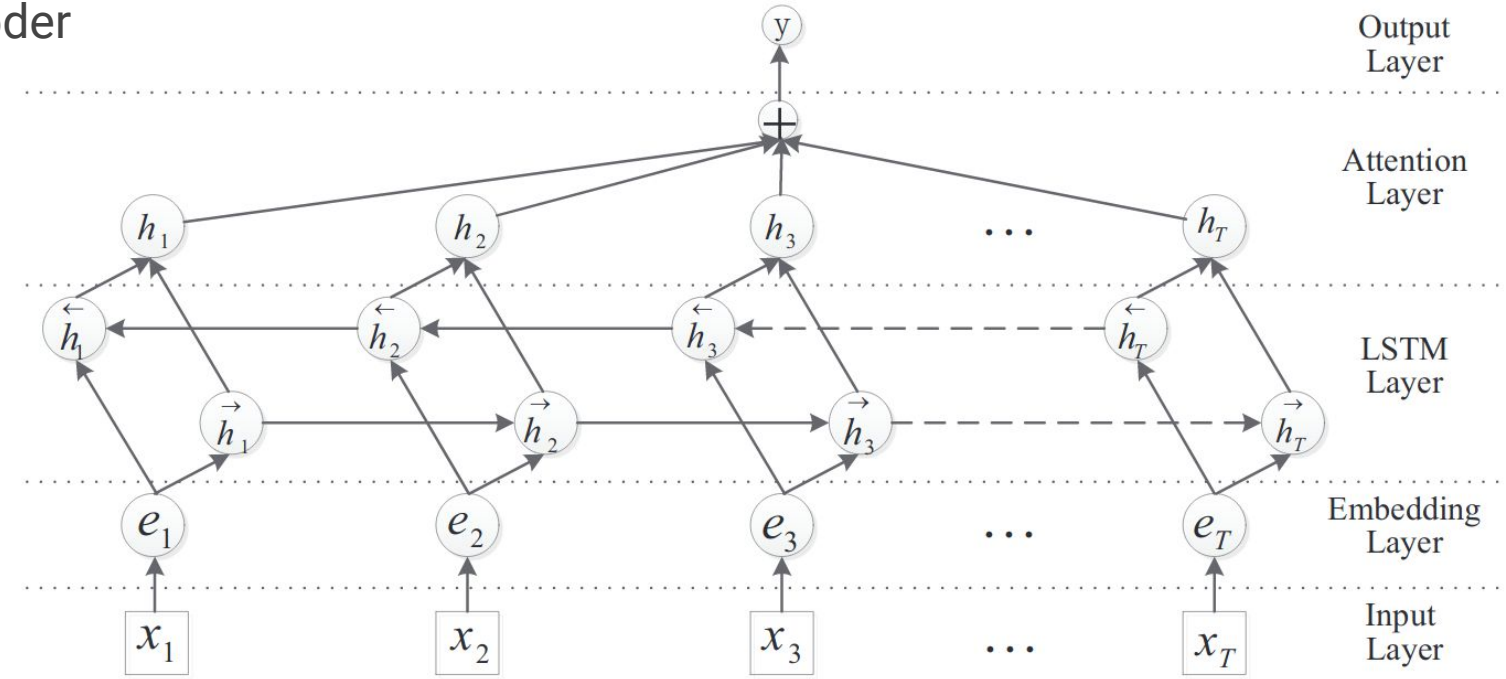


Encoder-Decoder



Attention

Encoder

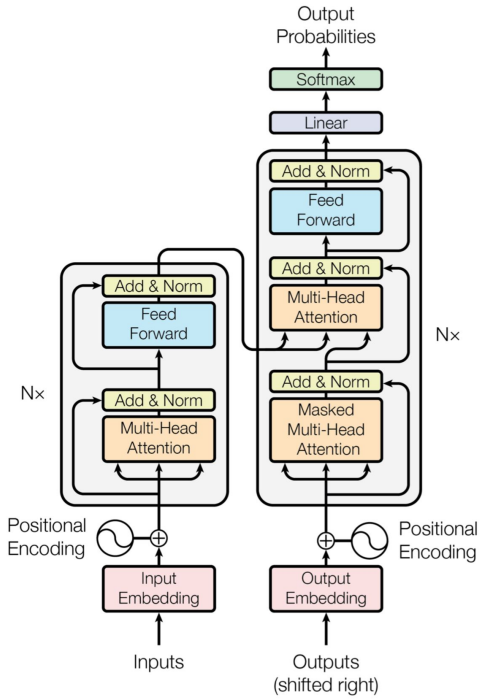


Transformer

Attention is all you need (Ashish Vaswani et al. 2017)

State-of-the-art in Machine Translation

Improvement on several NLP tasks over RNNs

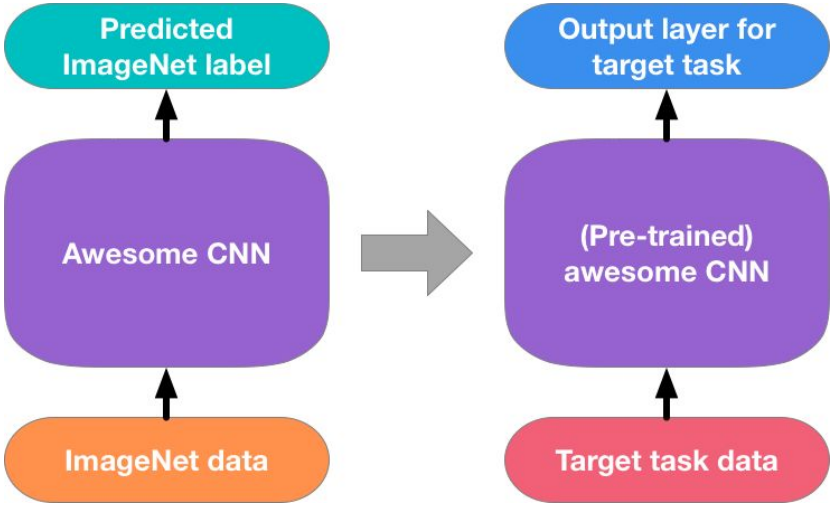


Contextualized Word Embeddings

Preliminary Concepts

Fine-tuning

Fine-tuning is the process in which parameters of a model must be adjusted very precisely in order to fit with certain observations.



Preliminary Concepts

Language Model

Language Model: Probabilistic model that is able to predict the next word in the sequence given the words that precede it.

Almost infinite training corpora

Typically trained using RNNs

Can be used as a feature extractor

Can go forward or backward

Word Embeddings

Types

Word Embeddings (one embedding per word)

- I'm a huge metal fan
- I went to a heavy metal concert
- This table is made of metal

Sense Embeddings (one embedding per sense)

- I'm a huge metal fan
- I went to a heavy metal concert
- This table is made of metal

Contextualized Word Embeddings (one embedding per word and context)

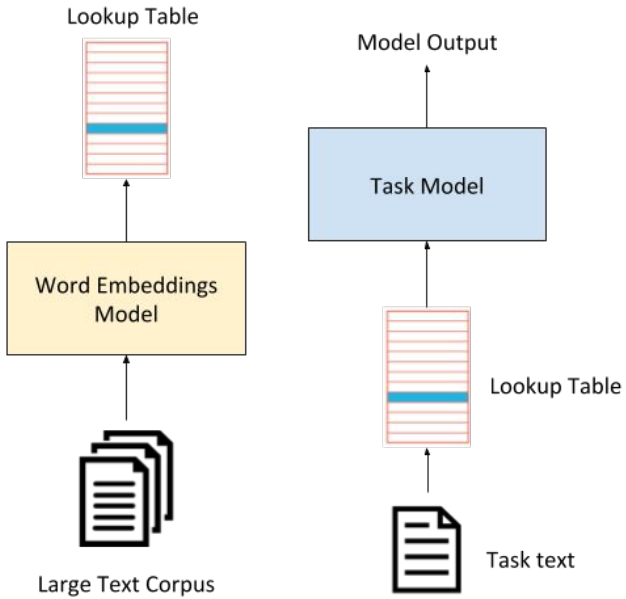
- I'm a huge metal fan
- I went to a heavy metal concert
- This table is made of metal

Word Embeddings

Word Embeddings

Single vector per word:

- First transfer learning approach in NLP,
- Pre-trained on external large corpora
- Used as features in different tasks and domains
- Popular approaches
  - word2vec (Mikolov et al., 2013)
  - Glove (Pennington et al., 2014)
  - fastText (Bojanowski et al., 2017)



Word Embeddings

Sense Embeddings

Several vectors per word (one vector per sense):

Popular approaches:

Nasari: Wikipedia entity embeddings

(José Camacho-Collados et al. 2016)

DeConf: WordNet sense embeddings

(M. T. Pilehvar et al. 2016)

...

metal

Word Embeddings

Contextualized embeddings

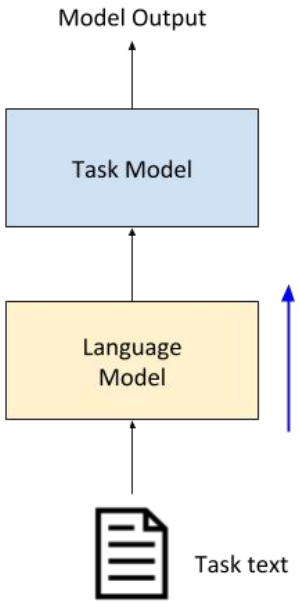
Several vectors per word (one vector per context):

Breakthrough in 2018! “ImageNet-like” momentum for NLP

Feature extractor model instead of a lookup table

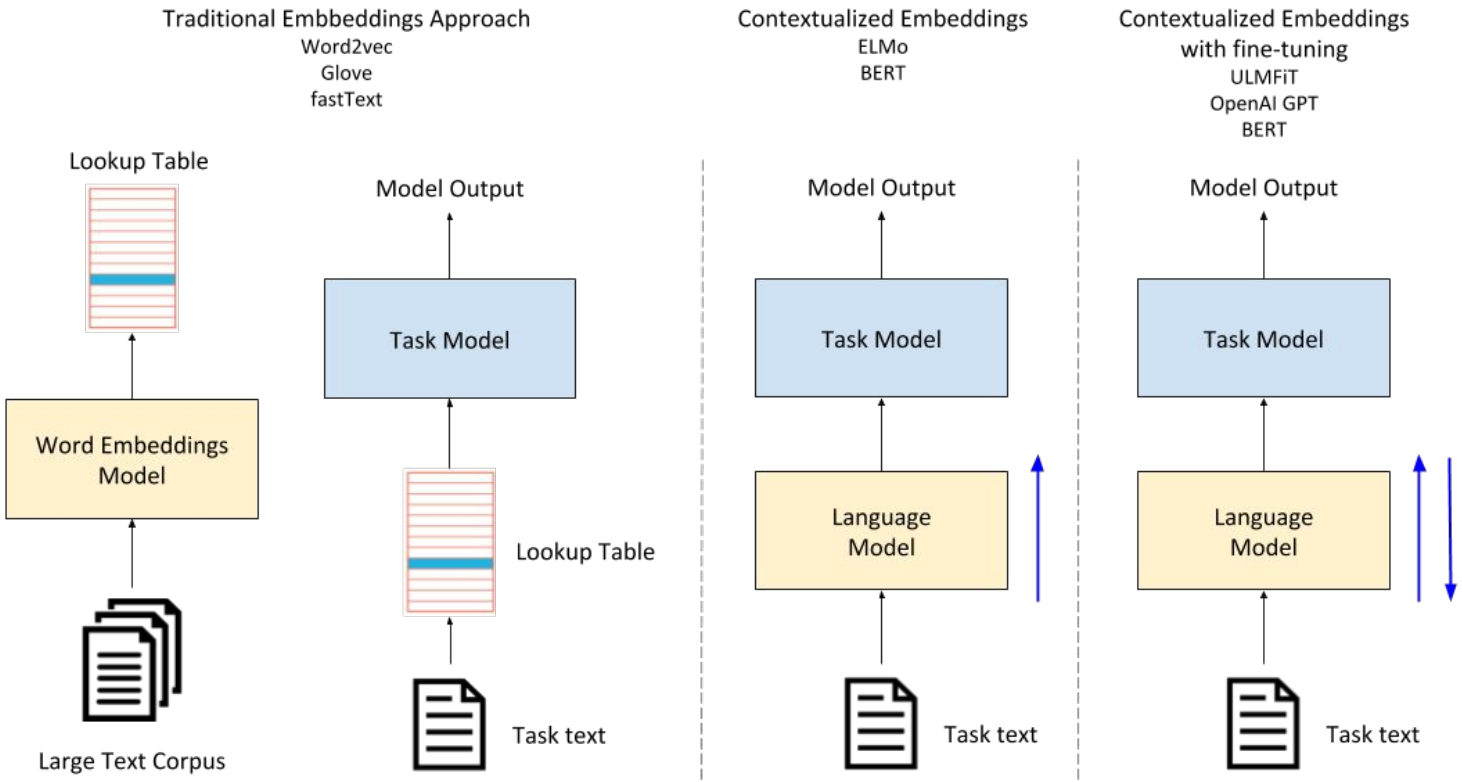
Popular contextualized embedding approaches:

- ELMo
- ULMFiT
- OpenAI GPT
- BERT



Word Embeddings

Summary



Further Reading

Interesting blog posts

https://jalammar.github.io/illustrated-bert/

http://jalammar.github.io/illustrated-transformer/

https://towardsdatascience.com/beyond-word-embeddings-part-2-word-vectors-nlp-modeling-from-bow-to-bert-4ebd4711d0ec

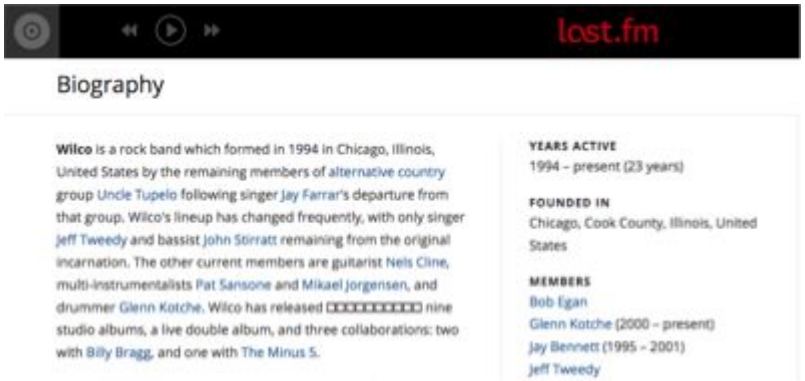
Multimodality

Multimodality

In large music collections information is multimodal:

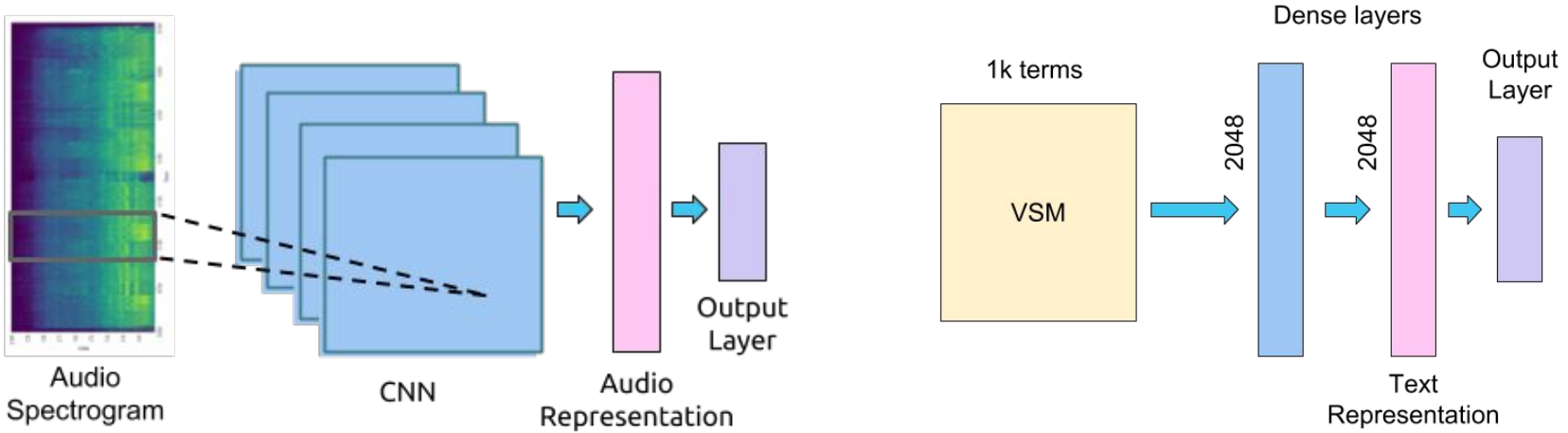
audio, images, texts, videos

How can we combine different modalities effectively?



Learn Data Representations

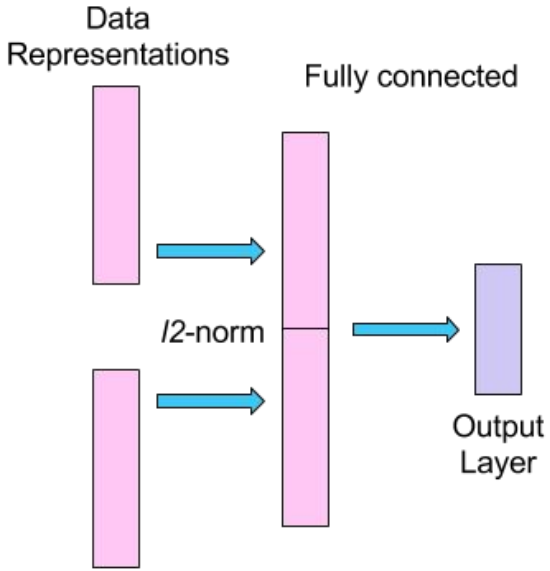
Train on the same task



Fusion of Data Representations

Extract internal representation from each modality

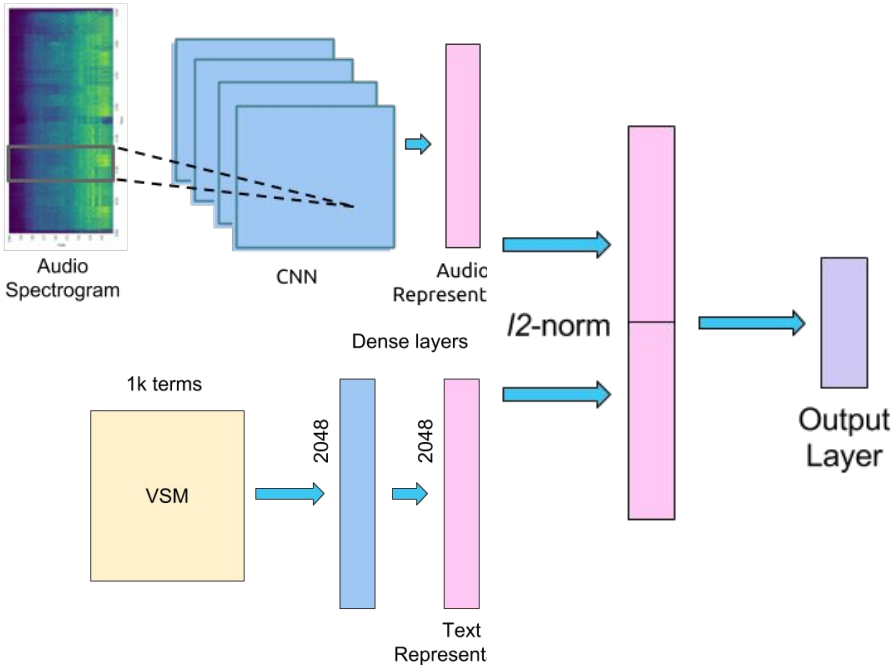
Concatenate and retrain on the same task



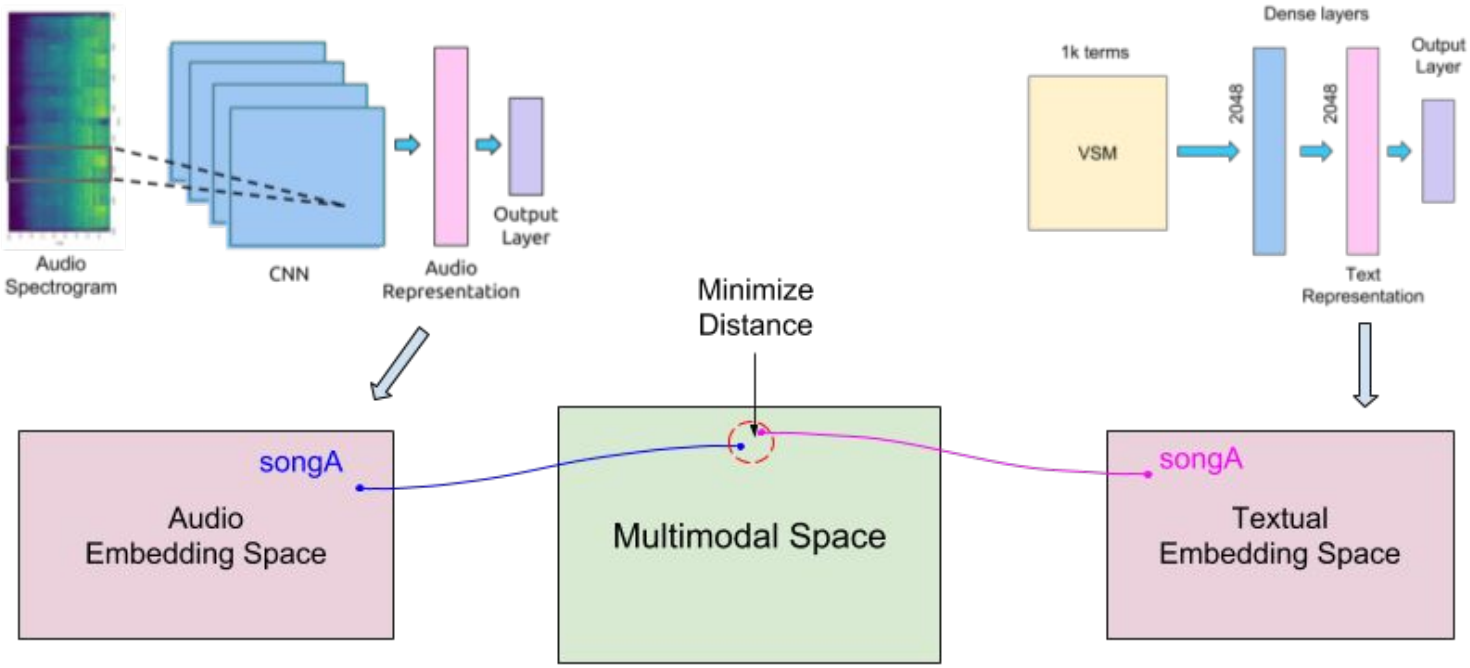
Fine-tuning of Multimodal Network

Pre-train modalities in isolation

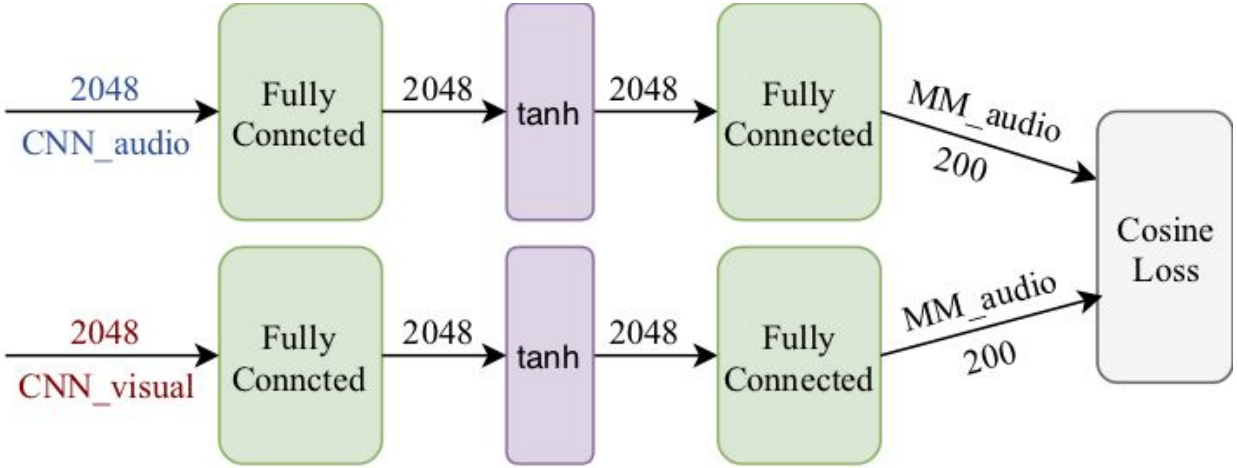
Multimodal Fusion with full networks



Multimodal Space



Multimodal Space



References

Oramas, S., Barbieri F., Nieto O., & Serra X. (2018). Multimodal Deep Learning for Music Genre Classification. Transactions of the International Society for Music Information Retrieval. 1(1), 4-21.

Oramas, S., Bogdanov D., & Porter A. (2018). MediaEval 2018 AcousticBrainz Genre Task: A baseline combining deep feature embeddings across datasets. MediaEval 2018 Workshop

Oramas, S., Nieto O., Sordo M., & Serra X. (2017). A Deep Multimodal Approach for Cold-start Music Recommendation. 2nd Workshop on Deep Learning for Recommender Systems, at RecSys 2017

Oramas, S., Nieto O., Barbieri F., & Serra X. (2017). Multi-label Music Genre Classification from Audio, Text and Images Using Deep Features. 18th International Society for Music Information Retrieval Conference (ISMIR 2017)

Conversational Systems

The rise of Conversational Assistants

Tech Chart of the Day

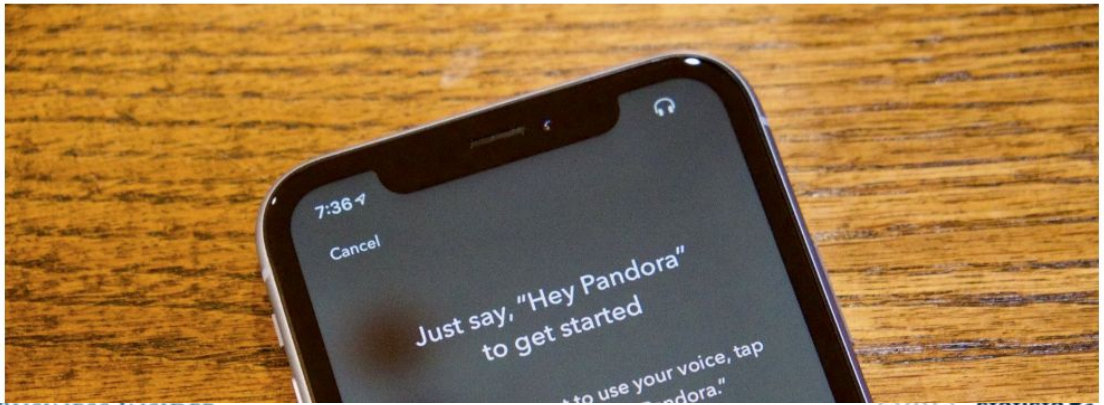
Pandora's new voice search feature knows what you want to hear

The secret sauce for Pandora's Voice Mode is personalized music, regardless of what you ask.

Nathan Ingraham
01.15.19 in AV

0
Comments

399
Shares



BUSINESS INSIDER

Sources: Experian, Creative Strategies

statista

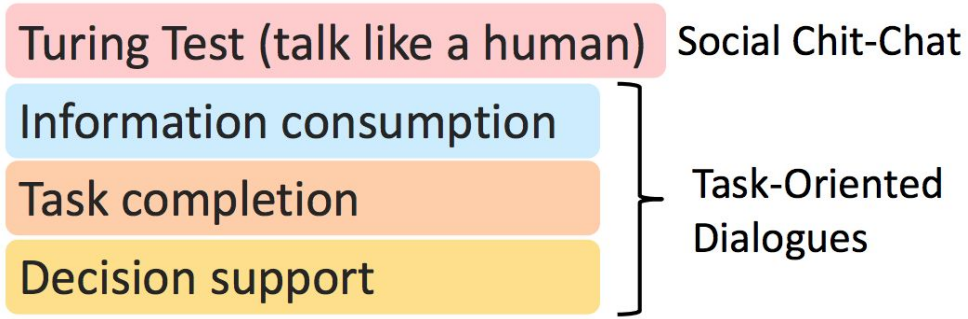
Task-oriented vs. Chit-Chat

"I want to chat"

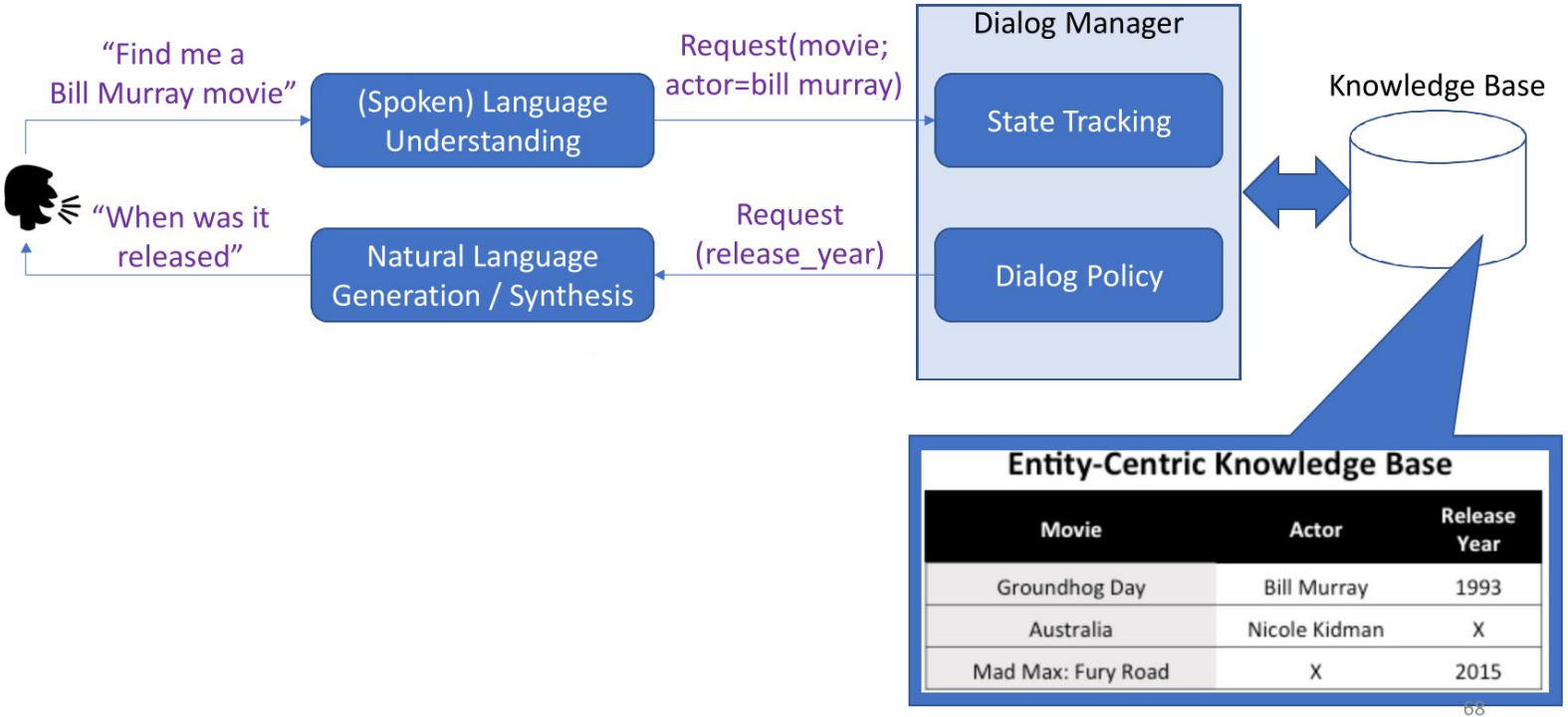
"I have a question"

"I need to get this done"

"What should I do?"



Task-Oriented Dialog Workflow



Task-Oriented Dialog Workflow

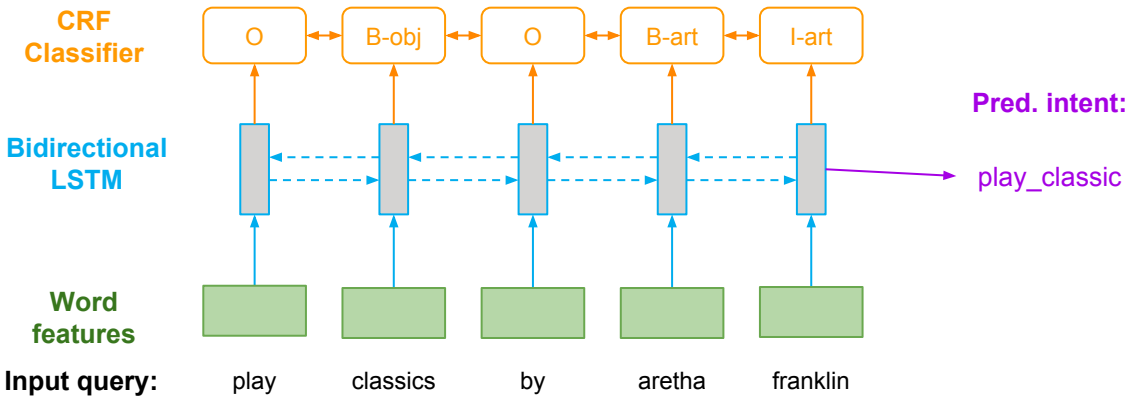
Spoken Language Understanding

Parse a transcribed oral query into slots and intents

"Play 90s rock with accordion"

Slots	Decade: 90s Genre: rock Instrument: accordion
Intent	play_music

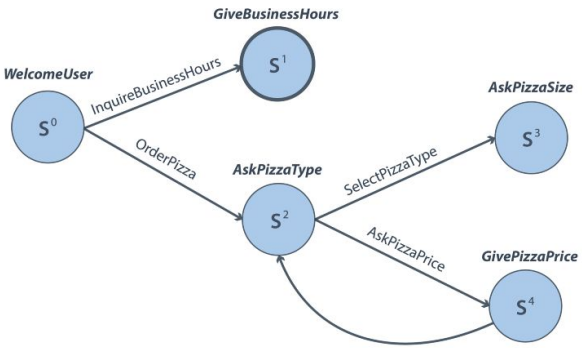
Stacked BiLSTM+CRF (Reimers and Gurevych, 2017)



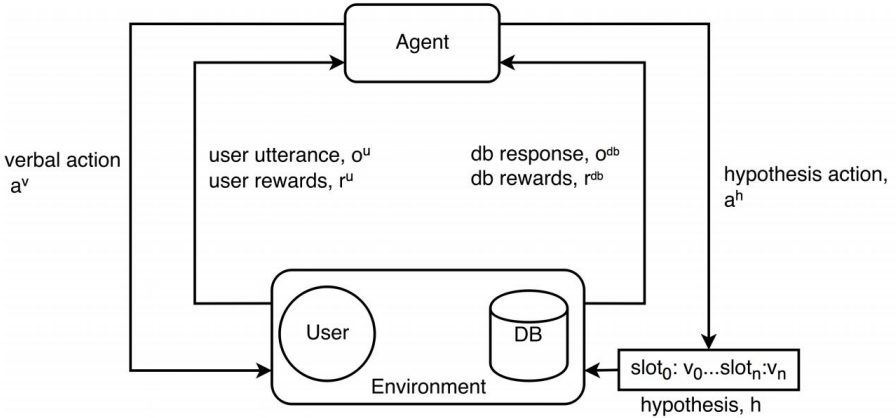
Task-Oriented Dialog Workflow

Dialog Policy

Finite State Machines



Reinforcement Learning (Zhano and Eskenazi, 2016)



Natural Language Generation

Translate a meaning representation into natural language utterances

Inform(artist=Metallica, genre=heavy_metal, artist_type=band)	“Metallica is a heavy metal band”

Template-based NLG

Inform(artist=$A, genre=$G, artist_type=$T) ⇒ “$A is a $G $T”

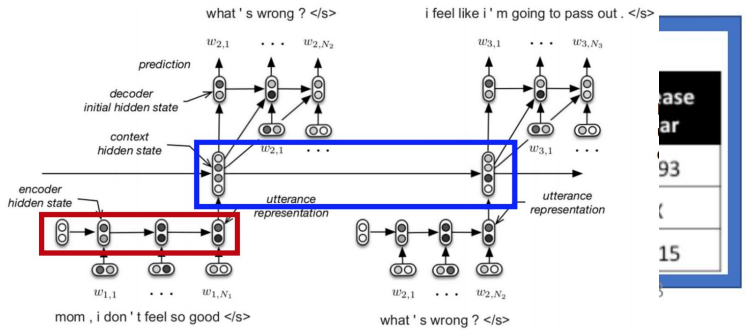
Generative models

RNN-based NLG (Wen+, 2015)

Data-driven E2E Dialog Systems

Hierarchical Encoder-Decoder (Serban+,2016)

Memory Networks (Bordes+, 2017)



Wanna dive more in depth?

Tutorials

- Neural approaches to Conversational AI [slides, arxiv]
- Deep learning for Conversational AI [slides]

Workshops

- SCAI: Search Oriented Conversational AI (ed. 2019 @ WWW2019 in SF!)
- CAIR: Conversational Approaches to IR

Blog posts

- High-density dialog management by b4.ai