Machine-Learning Semantics for Webscale NLP

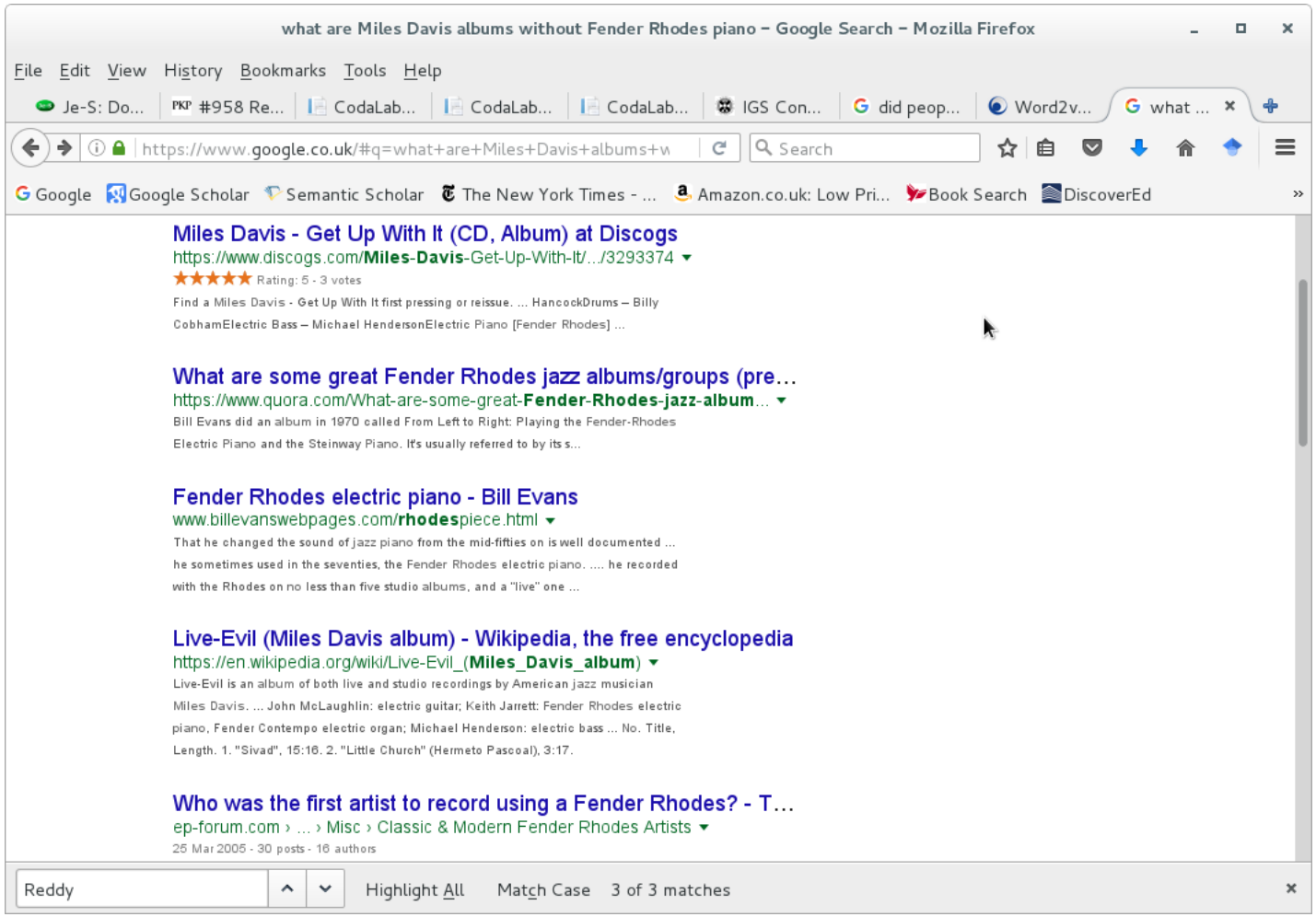
Mark Steedman

7th February 2017

School of informatics

The Problem of Content

- We have (somewhat) robust wide coverage parsers that work on the scale of Bn of words.They can read the web (and build logical forms) much faster than we can ourselves.
- So why can't we have them read the web for us, so that we can ask them questions like "What are recordings by Miles Davis without Fender Rhodes piano", and get a more helpful answer than the following?


what are Miles Davis albums without Fender Rhodes piano – Google Search – Mozilla Firefox
File Edit View History Bookmarks Tools Help
https://www.google.co.uk/#q=what+are+Miles+Davis+albums+w
Search
Google Google Scholar Semantic Scholar The New York Times - ... Amazon.co.uk: Low Pri... Book Search DiscoverEd
Miles Davis - Get Up With It (CD, Album) at Discogs
https://www.discogs.com/Miles-Davis-Get-Up-With-It/.../3293374
Rating: 5 - 3 votes
Find a Miles Davis - Get Up With It first pressing or reissue. ... HancockDrums – Billy CobhamElectric Bass – Michael HendersonElectric Piano [Fender Rhodes] ...
What are some great Fender Rhodes jazz albums/groups (pre...
https://www.quora.com/What-are-some-great-Fender-Rhodes-jazz-album...
Bill Evans did an album in 1970 called From Left to Right: Playing the Fender-Rhodes Electric Piano and the Steinway Piano. It's usually referred to by its s...
Fender Rhodes electric piano - Bill Evans
www.billevanswebpages.com/rhodespiece.html
That he changed the sound of jazz piano from the mid-fifties on is well documented ... he sometimes used in the seventies, the Fender Rhodes electric piano. .... he recorded with the Rhodes on no less than five studio albums, and a "live" one ...
Live-Evil (Miles Davis album) - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Live-Evil_(Miles_Davis_album)
Live-Evil is an album of both live and studio recordings by American jazz musician Miles Davis. ... John McLaughlin: electric guitar; Keith Jarrett: Fender Rhodes electric piano, Fender Contempo electric organ; Michael Henderson: electric bass ... No. Title, Length. 1. "Sivad", 15:16. 2. "Little Church" (Hermeto Pascoal), 3:17.
Who was the first artist to record using a Fender Rhodes? - T...
ep-forum.com › ... › Misc › Classic & Modern Fender Rhodes Artists
25 Mar 2005 - 30 posts - 16 authors
Reddy
Highlight All
Match Case
3 of 3 matches

Too Many Ways of Answering The Question

- The central problem of QA is that there are too many ways of asking and answering questions, and we have no idea of the semantics that relates them.
- Your Question: *Has Verizon bought Yahoo?*
- The Text:

1. Verizon purchased Yahoo. ("Yes")
2. Verizon's purchase of Yahoo ("Yes")
3. Verizon owns Yahoo ("Yes")
4. Verizon managed to buy Yahoo. ("Yes")
5. Verizon acquired every company. ("Yes")
6. Yahoo may be sold to Verizon. ("Maybe")
7. Verizon will buy Yahoo or Yazoo. ("Maybe not")
8. Verizon didn't take over Yahoo. ("No")

The Approach

- Use the semantic parsers to Machine-Read multiple relations over Named Entities in web text.
- Capture relations of Entailment over relations between NEs of the same types (Lewis and Steedman, 2013a,b, 2014; Lewis, 2015).
  - If you read somewhere that a person—say, Obama—was elected to an office—say, President—than you are highly likely to also read somwhere that that person ran for that office.
  - —but not the other way round
- Redefine the parser semantics in terms of entailments and paraphrases, and reparse and index the entire text for IR.

(There is another approach.)

Local Entailment Probabilities

- The typed named-entity technique is applied to (errorfully) estimate local probabilities of entailments using Weeds precision assymetric similarity (Weeds and Weir, 2003):

  a. $p(conquer\,x\,y \Rightarrow invade\,x\,y) = 0.9$

  b. $p(invade\,x\,y \Rightarrow attack\,x\,y) = 0.8$

  c. $p(conquer\,x\,y \Rightarrow attack\,x\,y) = 0.4$

  d. $p(bomb\,x\,y \Rightarrow attack\,x\,y) = 0.7$

  e. $p(bomb\,x\,y \Rightarrow conquer\,x\,y) = 0.2$

  (etc.)

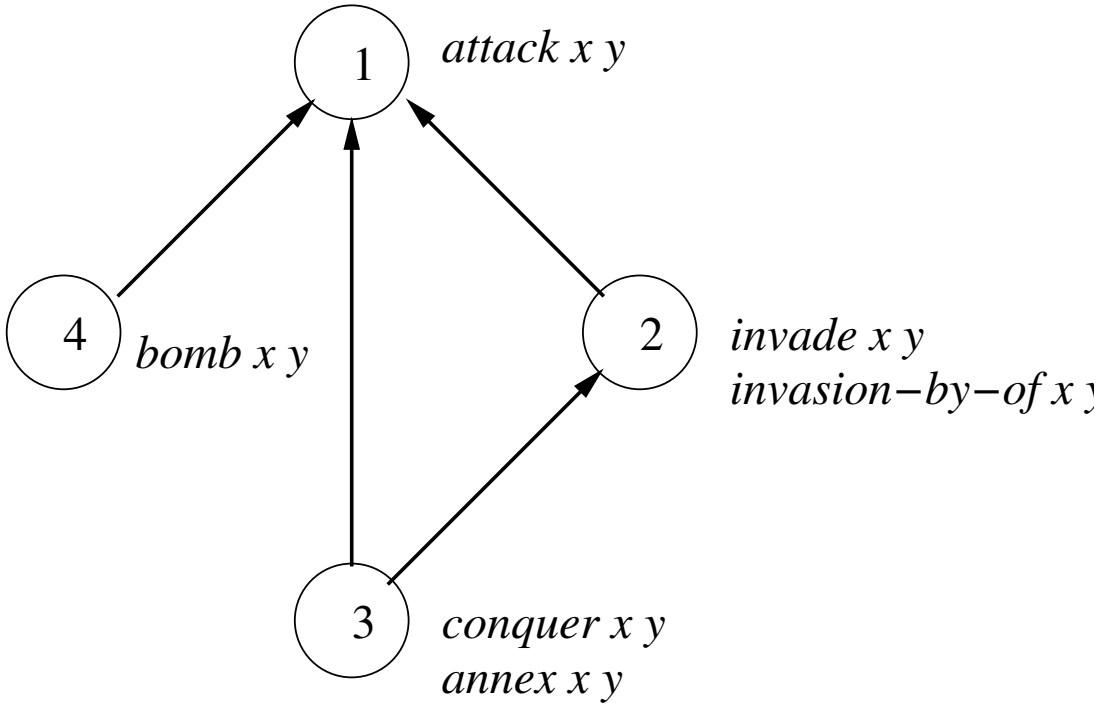
Global Entailments

- The local entailment probabilities are used to construct an entailment graph using integer linear programming with a prior $p = 0.25$ with the global constraint that the graph must be closed under transitivity (Berant *et al.*, 2015).

- Thus, (c) will be included despite low observed frequency, while other low frequency spurious local entailments will be excluded..

- Cliques within the entailment graphs are collapsed to a single paraphase cluster relation identifier.

Entailment graph



- A simple entailment graph for relations between countries.

Lexicon

- The new semantics obtained from the entailment graph

  attack := $(S\backslash NP)/NP : \lambda x \lambda y \lambda e.rel_1\, x\, y\, e$
  bomb := $(S\backslash NP)/NP : \lambda x \lambda y \lambda e.rel_1\, x\, y\, e \wedge rel_4\, x\, y\, e$
  invade := $(S\backslash NP)/NP : \lambda x \lambda y \lambda e.rel_1\, x\, y\, e \wedge rel_2\, x\, y\, e$
  conquer := $(S\backslash NP)/NP : \lambda x \lambda y \lambda e.rel_1\, x\, y\, e \wedge rel_2\, x\, y\, e \wedge rel_3\, x\, y\, e$
  annex := $(S\backslash NP)/NP : \lambda x \lambda y \lambda e.rel_1\, x\, y\, e \wedge rel_2\, x\, y\, e \wedge rel_3\, x\, y\, e$

- These logical forms support correct inference under negation, such as that *conquered* entails *attacked* and *didn't invade* entails *didn't conquer*

Results

- Examples:

Question	Answer	From Unseen Sentence:
What did Delta merge with?	Northwest	The 747 freighters came with Delta's acquisition of Northwest
What spoke with Hu Jintao?	Obama	Obama conveyed his respect for the Dalai Lama to China's president Hu Jintao during their first meeting
What arrived in Colorado?	Zazi	Zazi flew back to Colorado. . .
What ran for Congress?	Young	. . . Young was elected to Congress in 1972

- Full results in Lewis and Steedman (2013a)

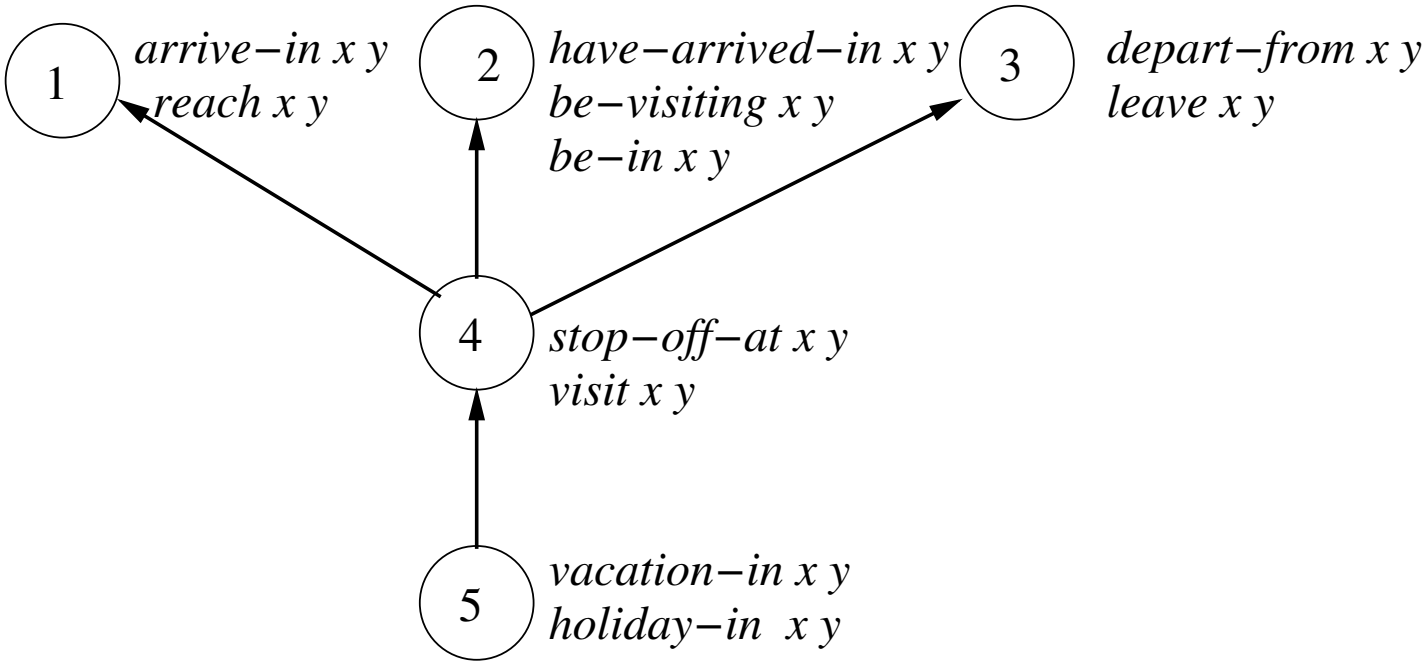
More Examples

Premise	Hypothesis	Answer
Obama want to boost the defense budget	Obama increase the defense budget	False
The thieves make off with TVs	The thieves manage to steal TVs	True
My son be terrified of him	My son have a fear of him	True

Multilingual Example

Source:	Le Princess Elizabeth arrive à Dunkerque le 3 août 1999
SMT 1-best:	The Princess Elizabeth is to manage to Dunkirk on 3 August 1999.
Reranked 1-best:	The Princess Elizabeth arrives at Dunkirk on 3 August 1999.

The Next Step: Generalize to Aspectual Semantics



- A simple entailment graph for relations over events does not capture relations of causation and temporal sequence.

Learning from Timestamped Data

- One source of information concerning these hidden relations is timestamped news, of the kind available in the University of Washington NEWSSPIKE corpus of 0.5M newswire articles (Zhang and Weld, 2013).
- In such data, we find that statements that so-and-so *is visiting*, *is in* and the perfect *has arrived in* such and such a place, occur in stories with the same datestamp, whereas *is arriving*, *is on her way to*, occur in preceding stories, while *has left, is on her way back from, returned*, etc. occur in later ones.
- This information provides a basis for inference that *visiting* entails *being in*, that the latter is the consequent state of *arriving*, and that *arrival* and *departure* coincide with the beginning and end of the progressive state of *visiting*.

Needs new datasets for evaluation.

Machine Reading into Semantic Net

- We would like to interrogate huge databases such as the Google knowledge graph, a.k.a. Semantic Nets (Reddy *et al.*, 2014)
- There is a mismatch between the semantics delivered by parsers and the language of the knowledge graph.
- So lets build our own knowledge graph using the clustered entailment semantics of the parser, so that we can query it directly in natural language.

This is a potentially a much bigger graph than the Knowledge Graph.

- We will need techniques to limit exponential growth in the costs of loading and interrogating this graph.
- Pilot experiments by Harrington and Clark (2009); Lao *et al.* (2012) suggest this can be done by spreading activation (Collins and Loftus, 1975).

References

Berant, Jonathan, Alon, Noga, Dagan, Ido, and Goldberger, Jacob, 2015. "Efficient Global Learning of Entailment Graphs." *Computational Linguistics* 42:221–263.

Collins, Allan and Loftus, Elizabeth, 1975. "A Spreading Activation Theory of Semantic Processing." *Psychological Review* 82:407–428.

Harrington, Brian and Clark, Stephen, 2009. "ASKNet: Creating and Evaluating Large Scale Integrated Semantic Networks." *International Journal of Semantic Computing* 2:343–364.

Lao, Ni, Subramanya, Amarnag, Pereira, Fernando, and Cohen, William, 2012. "Reading the Web with Learned Syntactic-Semantic Inference Rules." In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. ACL, 1017–1026.

Lewis, Mike, 2015. *Combined Distributional and Logical Semantics*. Ph.D. thesis, University of Edinburgh.

Lewis, Mike and Steedman, Mark, 2013a. "Combined Distributional and Logical Semantics." *Transactions of the Association for Computational Linguistics* 1:179–192.

Lewis, Mike and Steedman, Mark, 2013b. "Unsupervised Induction of Cross-Lingual Semantic Relations." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. ACL, 681–692.

Lewis, Mike and Steedman, Mark, 2014. "Combining Formal and Distributional Models of Temporal and Intensional Semantics." In *Proceedings of the ACL Workshop on Semantic Parsing*. Baltimore, MD: ACL, 28–32. Google Exceptional Submission Award.

Reddy, Siva, Lapata, Mirella, and Steedman, Mark, 2014. "Large-scale Semantic Parsing without Question-Answer Pairs." *Transactions of the Association for Computational Linguistics* 2:377–392.

Steedman, Mark, 2000. *The Syntactic Process*. Cambridge, MA: MIT Press.

Weeds, Julie and Weir, David, 2003. "A General Framework for Distributional Similarity." In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. ACL, 81–88.

Zhang, Congle and Weld, Daniel, 2013. "Harvesting Parallel News Streams to

Generate Paraphrases of Event Relations." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Seattle: ACL, 1776–1786.