The University of Manchester

Effective Semantics for Engineering NLP Systems

André Freitas

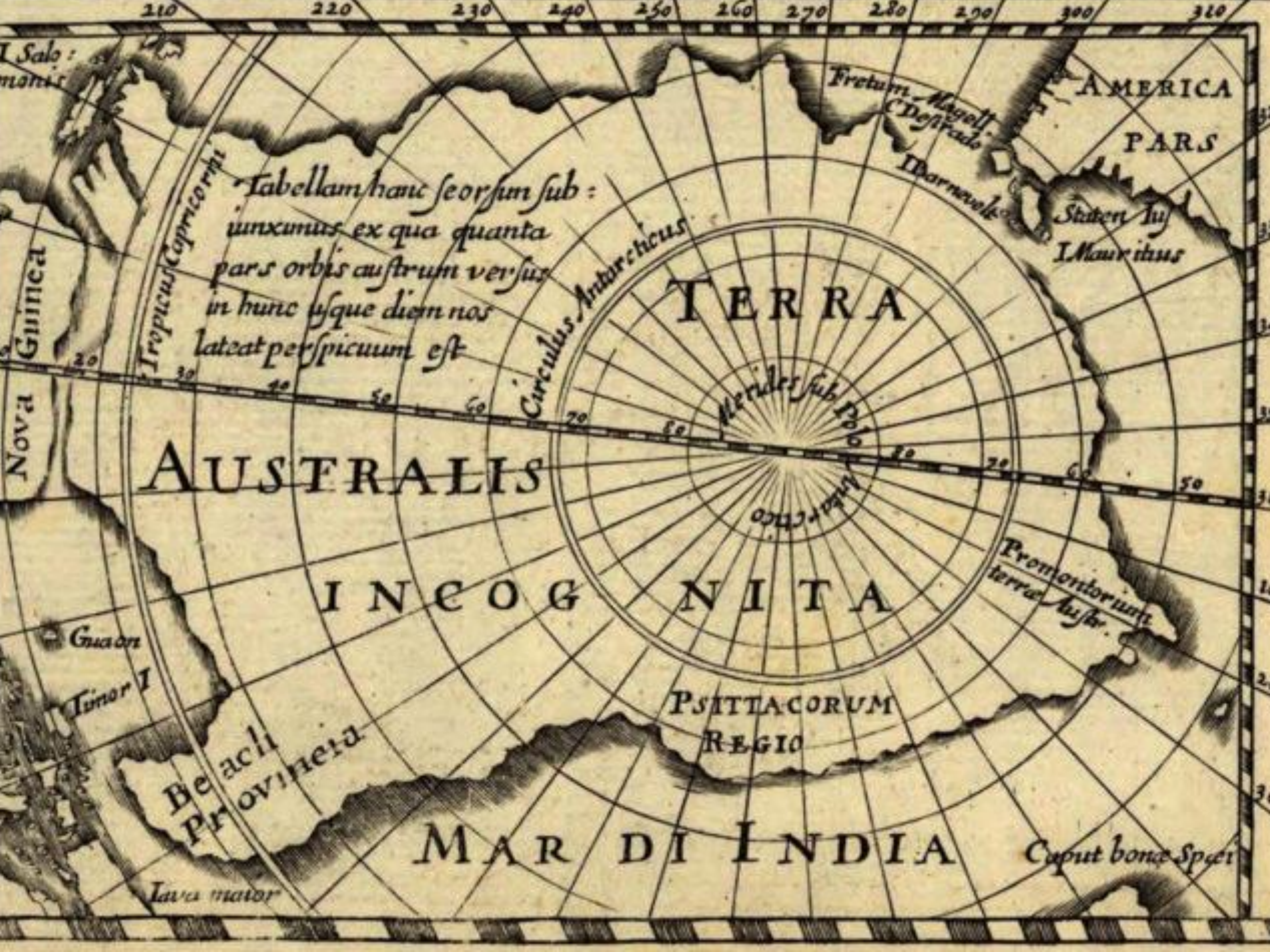
Lancaster, May 2018

Goals of this Talk

Provide a synthesis of the emerging representation trends behind NLP systems.

Shift in perspective:

- Effective engineering (task driven, scalable) instead of sound formalism.
- Best-effort representation.


210
220
230
240
250
260
270
280
290
300
310
I Salo: monis
Fretum Magell
C Destrado
I Barnevelt
AMERICA
PARS
Staten Ius
I Mauritius
Tabellam hanc seorsum sub:
iunximus ex qua quanta
pars orbis austrum versus
in hunc usque diem nos
lateat perspicuum est
Tropicus Capricorni
Circulus Antarcticus
Nova Guinea
TERRA
AUSTRALIS
INCOG NITA
Meridies sub Polo Antarctico
Promontorium terra Austr.
Timor I
PSITTACORUM
REGIO
Beach
Provincia
MAR DI INDIA
Caput bone Spei
Iava maior

Outline

- Knowledge Graphs (Frege revisited)
- Information Extraction & Text Classification
- Distributional Semantic Models
- Knowledge Graphs & Distributional Semantics
  - (Distributional-Relational Models)
- Applications of DRMs
  - KG Completion
  - Semantic Parsing
  - Natural Language Inference

“On our best behaviour”

Levesque, 2013

“We need to return to our roots in Knowledge Representation and Reasoning *for* language and *from* language.”

“We should not treat English text as a monolithic source of information.”

“Instead, we should carefully study how simple knowledge bases might be used to make sense of the simple language needed to build slightly more complex knowledge bases…”

MANCHESTER 1824
The University of Manchester

Knowledge Graphs (Frege Revisited)

Some Perspectives on "What"

"*The **Knowledge Graph** is a knowledge base used by Google to enhance its search engine's search results.*"

"A Knowledge graph (i) mainly describes real world entities and interrelations, organized in a graph (ii) defines possible classes and relations of entities in a schema (iii) allows potentially interrelating arbitrary entities with each other…" [Paulheim H.]

"We define a Knowledge Graph as an RDF graph consists of a set of RDF triples where each RDF triple (s,p,o)…." *[Pujara J. al al.]*

Some Perspectives on “What”

Dan Bennett, TR

- Open world representation of information.
- Every entry point is equal cost.
- Underpin Cortana, Google Assistant, Siri, Alexa.
- Typically (but doesn’t have to be) expressed in RDF.
- No longer a solution in search of a problem!

Some Perspectives on "Why"

- "Knowledge is Power" Hypothesis (the Knowledge Principle): "If a program is to perform a complex task well, **it must know a great deal about the world** in which it operates."

- The Breadth Hypothesis: "To behave intelligently in unexpected situations, an agent must be capable of falling back on **increasingly general knowledge**."

Some Perspectives on "Why"

- We're surrounded by **entities**, which are connected by **relations.**
- We need to store them somehow, e.g., using a **DB** or a **graph.**
- **Graphs** can be processed **efficiently** and offer a convenient **abstraction.**

Some Perspectives on “Why”

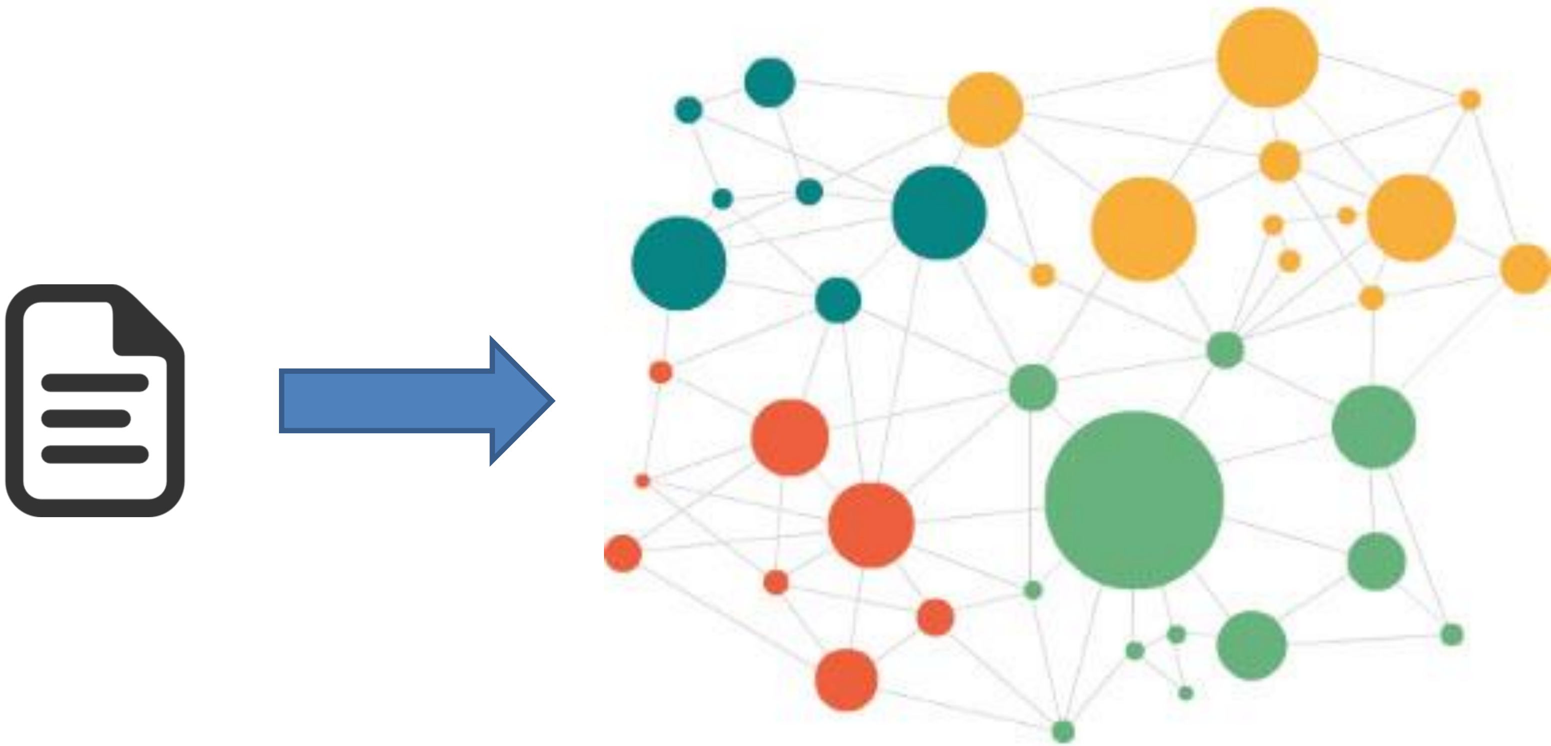
- Knowledge models such as Linked Data and many problems in machine learning have a natural representation as relational data.

- Relations between entities are often more important for a prediction task than attributes.

- For instance, can be easier to predict the party of a vice-president from the party of his president than from his attributes.

[Koopman, 2010]

MANCHESTER 1824
The University of Manchester



Building Knowledge Graphs

Open Information Extraction

- Extracting **unstructured** facts from text.
- **TextRunner** [Banko et al., IJCAI ’07], **WOE** [Wu & Weld, ACL ‘10].
- **ReVerb** [Fader et al., EMNLP ‘11].
- **OLLIE** [Mausam et al., EMNLP ‘12].
- **OpenIE** [Mausam et al., IJCAI ‘16].
- **Graphene** [Niklaus et al, COLING 17].

MANCHESTER 1824
The University of Manchester

Graphene

- Captures contextual relations.
- Extends the default Open IE representation in order to capture inter-proposition relationships.
- Include rhetorical relations.

Cetto et al., Creating a Hierarchy of Semantically-Linked Propositions in Open Information Extraction, COLING (2018).

Niklaus et al., A Sentence Simplification System for Improving Relation Extraction, COLING (2017)

Transformation Stage

non-restrictive relative clauses e.g. "*The city's top tourist attraction was the Notre Dame Cathedral, which welcomed 14 million visitors in 2013.*"

non-restrictive appositive phrases e.g. "*He plays basketball, a sport he participated in as a member of his high school's varsity team.*"

restrictive appositive phrases e.g. "*He met with former British Prime Minister Tony Blair.*"

participial phrases offset by commas e.g. "*The deal, titled Joint Comprehensive Plan of Action, saw the removal of sanctions.*"

adjective and adverb phrases delimited by punctuation e.g. "*Overall, the economy expanded at a rate of 2.9 percent in 2010.*"

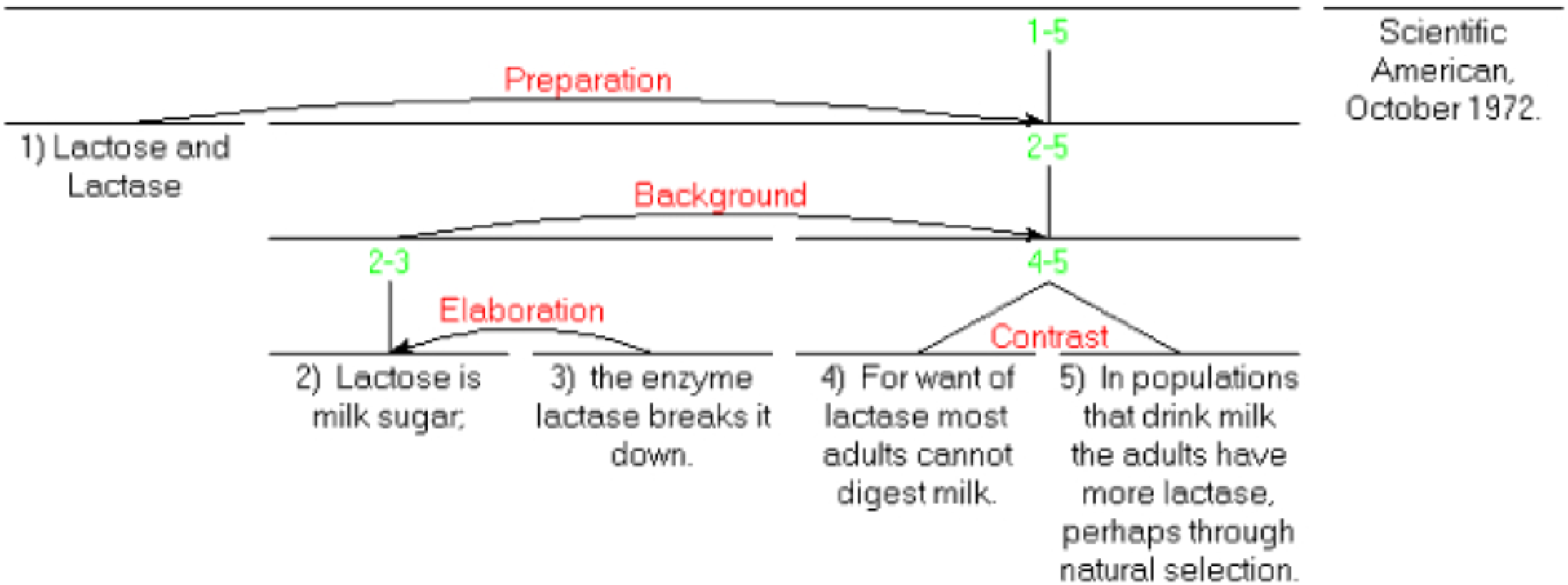
particular prepositional phrases e.g. "*In December 2008 and in 2012, Time magazine named Obama as its Person of the Year.*"

lead noun phrases e.g. "*Six weeks later, Alan Keyes accepted the Republican nomination.*"

intra-sentential attributions e.g. "*He said that both movements seek to bring justice and equal rights to historically persecuted peoples.*"

parentheticals e.g. "*He signed the reauthorization of the State Children's Health Insurance Program (SCHIP).*"

Rhetorical Relations



Extracting Rhetorical Relations

Rhetorical Relation / Inverse Rel.	Coordinations Core Span	Following Core Span
Unknown-Coordination / Unknown-Coordination	a syntactically coordinated span (default)	a syntactically coordinated span (default)
Contrast / Contrast	one alternative	the other alternative
Cause / Result	a situation	another situation which causes that one
Result / Cause	a situation	another situation which is caused by that one
List / List	a listed element	the next listed element
Disjunction / Disjunction	a listed element	a listed, alternative element
Temporal-After / Temporal-Before (Sequence)	a situation	a situation that occurs after that
Temporal-Before / Temporal-After (Inverted-Sequence)	a situation	a situation that occurs before that

Extracting Rhetorical Relations

	Subordinations	
Rhetorical Relation	Core Span	Context Span
Unknown Subordination	the syntactically superordinated span (default)	the syntactically subordinated span (default)
Attribution	the reported message	the source of the attribution
Background	text whose understanding is being facilitated	text for facilitating understanding
Cause	a situation	another situation which causes that one
Result	a situation	another situation which is caused by that one
Condition	action or situation whose occurrence results from the occurrence of the conditioning situation	conditioning situation
Elaboration	basic information	additional information
Purpose	an intended action	the intent behind the situation
Temporal-After	a situation	a situation that occurs after that
Temporal-Before	a situation	a situation that occurs before that

MANCHESTER 1824
The University of Manchester

Classified Contextual Relations

Relation	Core	Context
Unknown	a situation (default)	contextual information (default)
Noun Based	a situation	additional information about entities that are mentioned in the situation
Spatial	a situation	spatial information that describes where the situation took place
Temporal	a situation	temporal information that describes when the situation happened

Clausal & Phrasal Disembedding

input: NL text

"A few hours later, Matthias Goerne, a German baritone, offered an all-German program at the Frick Collection."

↓

syntax-based sentence simplification

↓

- *core sentence*: Matthias Goerne offered an all-German program.
- *context sentence*: Matthias Goerne was a German baritone.
- *context sentence*: This was a few hours later.
- *context sentence*: This was at the Frick Collection.

↓

relation extraction (using the Open IE system from UW)

↓

output: extractions (in JSON format)

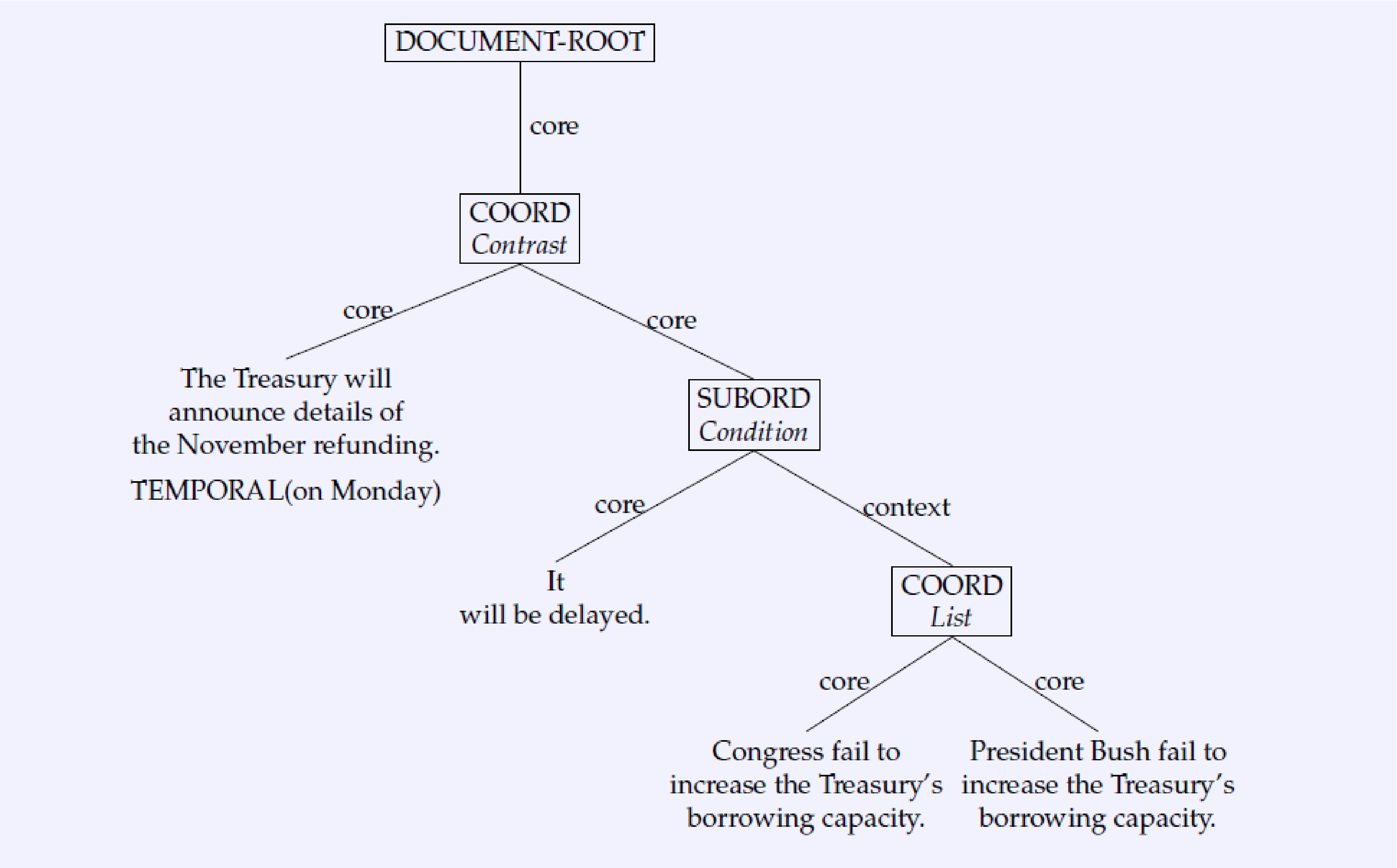
- *core fact*: offered (Matthias Goerne; an all-German program)
- *context 1*: was (Matthias Goerne; a German baritone)
- *context 2*: was (CORE FACT; at the Frick Collection)
- *context 3*: was (CORE FACT; a few hours later)

The University of Manchester

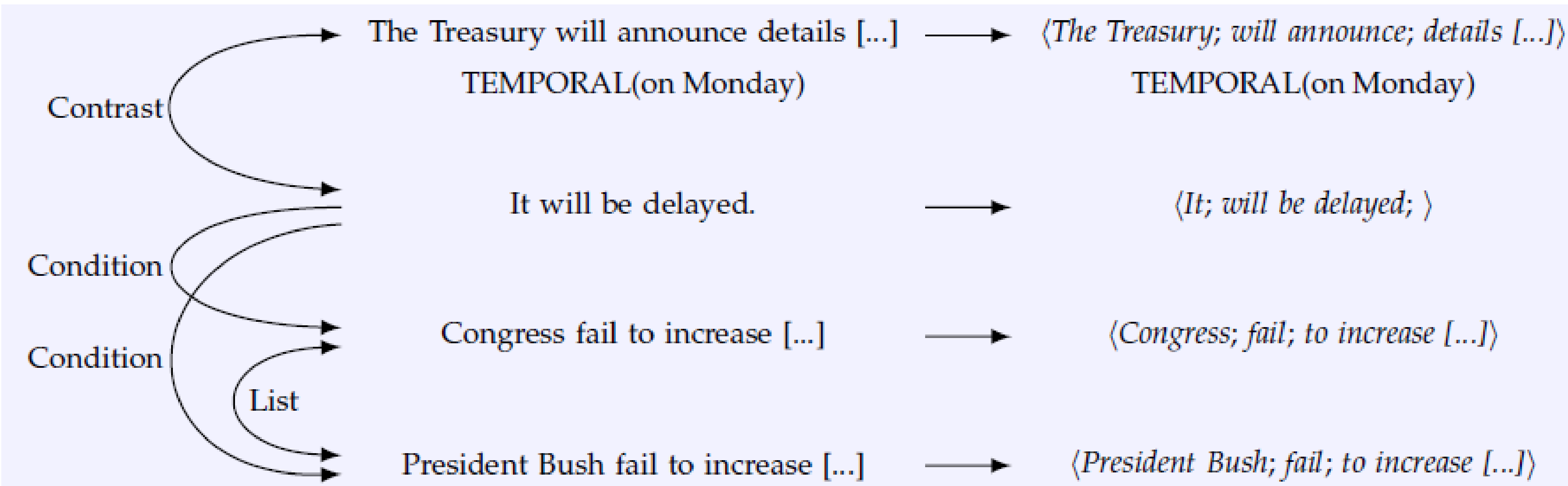
Input Document

[...] Although the Treasury will announce details of the November refunding on Monday, it will be delayed if Congress and President Bush fail to increase the Treasury's borrowing capacity. [...]

Transformation Stage



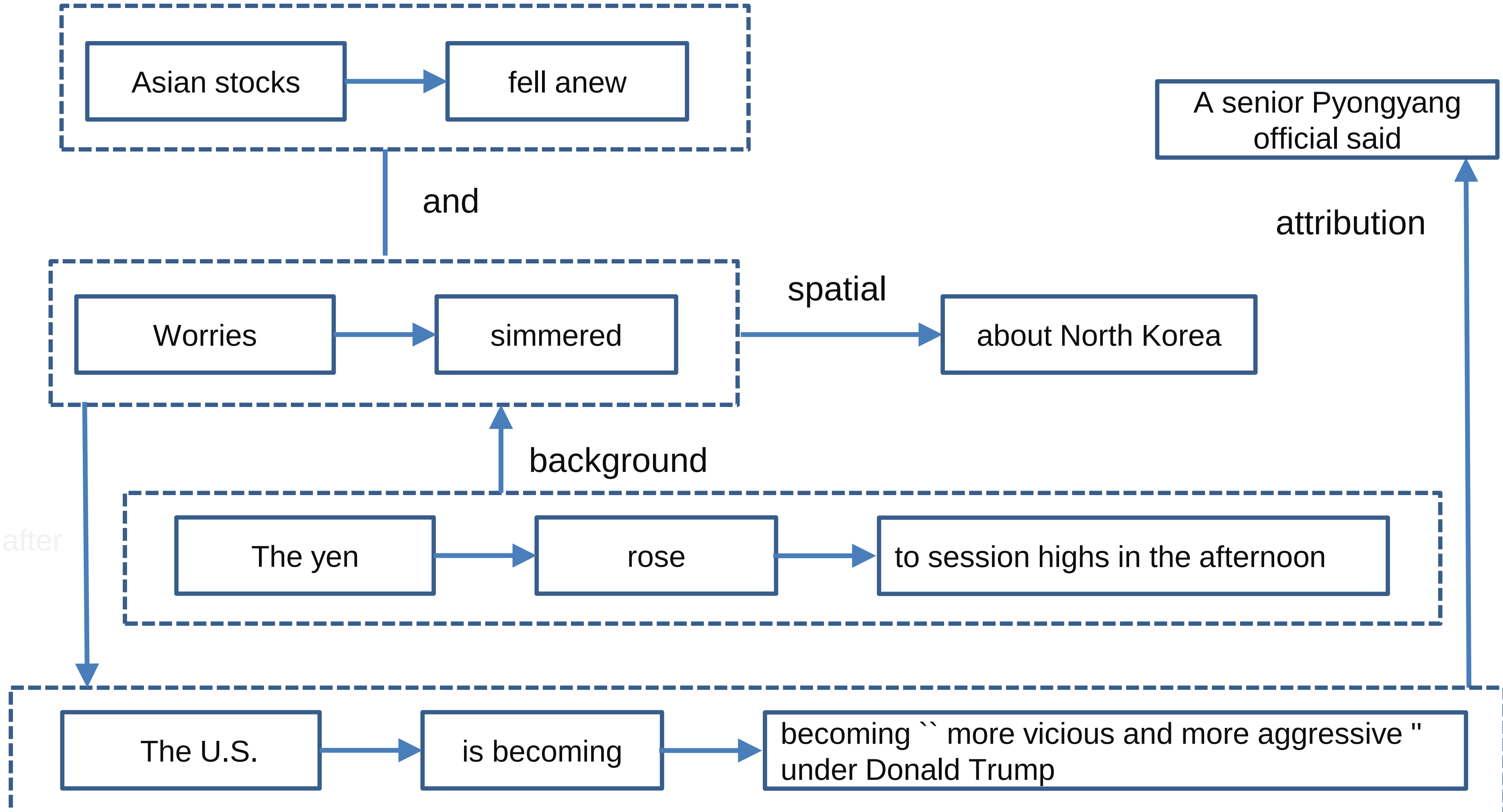
Relation Extraction



Output

```
#1    0    the Treasury    will announce    details [...]
    S:TEMPORAL    on Monday
    L:CONTRAST    #2

#2    0    it    will be delayed
    L:CONTRAST     #1
    L:CONDITION    #3
    L:CONDITION    #4

#3    1    Congress    fail    to increase [...]
    L:LIST    #4

#4    1    president Bush    fail    to increase [...]
    L:LIST    #3
```

Asian stocks fell anew and the yen rose to session highs in the afternoon as worries about North Korea simmered, after a senior Pyongyang official said the U.S. is becoming ``more vicious and more aggressive" under President Donald Trump .



He nominated Sonia Sotomayor on May 26, 2009 to replace David Souter; she was confirmed on August 6, 2009, becoming the first Supreme Court Justice of Hispanic descent.

OLLIE:

(1)	she	was confirmed on	August 6, 2009
(2)	He	nominated Sonia Sotomayor on	May 26
(3)	He	nominated Sonia Sotomayor	2009
(4)	He	nominated 2009 on	May 26
(5)	Sonia Sotomayor	be nominated 2009 on	May 26
(6)	He	nominated 2009	Sonia Sotomayor
(7)	2009	be nominated Sonia Sotomayor on	May 26

ClausIE:

(8)	He	nominated	Sonia Sotomayor on May 26 2009 to replace David Souter
(9)	she	was confirmed	on August 6 2009 becoming the first Supreme Court Justice of Hispanic descent
(10)	she	was confirmed	becoming the first Supreme Court Justice of Hispanic descent

Graphene:

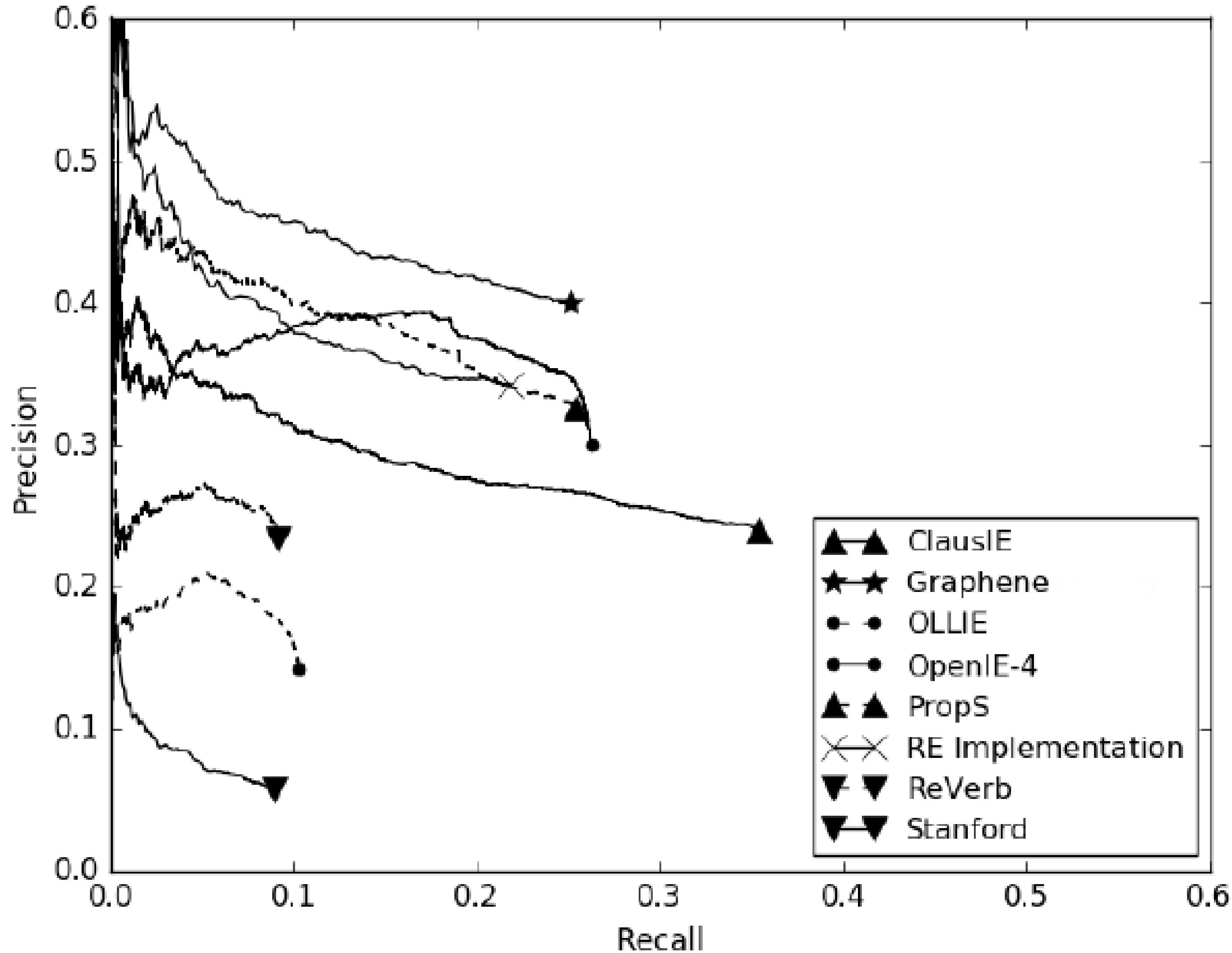
(11)	**#1**	**0**	**he**	**nominated**	**Sonia Sotomayor**
("a)		**S:PURPOSE**		**to replace David Souter**	
("b)		**S:TEMPORAL**		**on May 26, 2009**	
(12)	**#2**	**0**	**she**	**was confirmed**	
("a)		**S:TEMPORAL**		**on August 6, 2009**	
(13)	**#3**	**0**	**she**	**was becoming**	**the first Supreme Court Justice of Hispanic descent**

```
Although the Treasury will announce details of the November refunding on Monday, the funding
will be delayed if Congress and President Bush fail to increase the Treasury's borrowing ca-
pacity.

OLLIE:
(14) the Treasury                  will announce     details of the November refunding
(15) Congress and President Bush   fail to increase  the Treasury's borrowing capacity

ClausIE:
(16) the Treasury   will announce     details of the November refunding on Monday
(17) the Treasury   will announce     details of the November refunding
(18) the funding    will be delayed   if Congress and President Bush fail to increase the
                                            Treasury 's [...]
(19) the funding    will be delayed   if Congress and President Bush fail to increase the
                                            Treasury 's [...] Although the Treasury will
                                            announce details of the November [...]
(20) Congress and President Bush   fail   to increase the Treasury 's borrowing capacity
(21) the Treasury   has   borrowing capacity
```

Graphene:

```
(22) #1    0    the Treasury    will announce    details of the November refunding
("a)       S:TEMPORAL    on Monday
("b)       L:CONTRAST     #2
(23) #2    0    the funding    will be delayed
("a)       L:CONTRAST     #1
("b)       L:CONDITION    #3
("c)       L:CONDITION    #4
(24) #3    1    Congress    fail    to increase the Treasury 's borrowing capacity
(25) #4    1    president Bush    fail    to increase the Treasury 's borrowing capacity
```

MANCHESTER 1824
The University of Manchester



What to expect? (Wikipedia & Newswire)

Precision:

System	# Extractions		Accuracy			Minimality			Limitation to core information	Context allo-cation	Accuracy & Minimality	All
	T	C	T	T & C	C	T	T & C	C	T	C	T & C	T & C
Graphene	182	**192**	**0.82**	**0.82**	**0.81**	**0.74**	**0.79**	**0.84**	**0.87**	**0.91**	**0.66**	**0.61**
OLLIE	438	12	0.42	0.43	0.58	0.64	0.64	0.58	0.50	0.58	0.31	0.19
ClausIE	535	0	0.52	0.52	—	0.57	0.57	—	0.36	—	0.24	0.14
Stanford OIE	**1073**	0	0.26	0.26	—	0.63	0.63	—	0.53	—	0.19	0.13

Recall:

System	accurate		accurate & minimal	
	P	P & A	P	P & A
Graphene	0.78	0.69	0.80	0.77
OLLIE	0.69	0.55	0.70	0.63
ClausIE	**0.83**	**0.83**	**0.84**	**0.83**
Stanford OIE	0.62	0.46	0.63	0.54

Improving Open Relation Extraction using Clausal and Phrasal Disembedding, Under Review, (2017)

MANCHESTER 1824
The University of Manchester

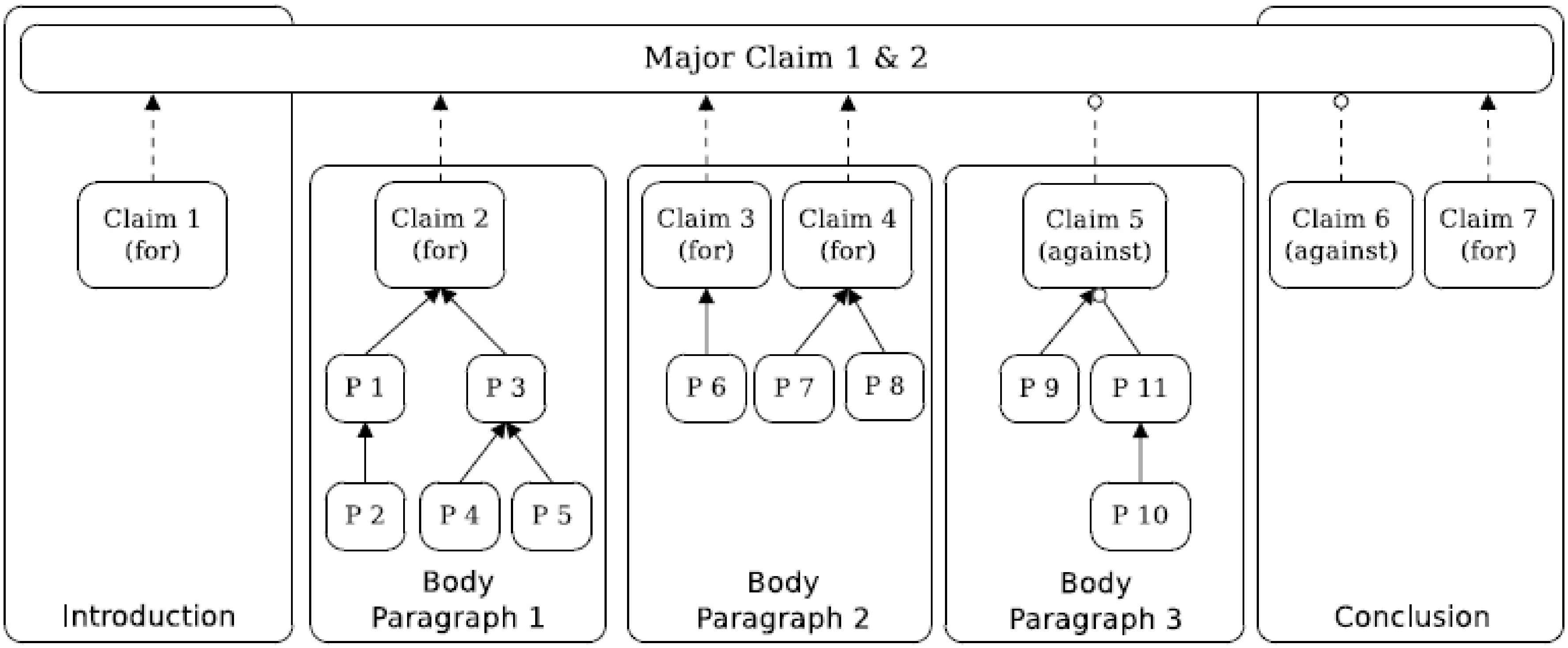
Software: Extracting Knowledge Graphs from Text

Graphene

https://github.com/Lambda-3/Graphene

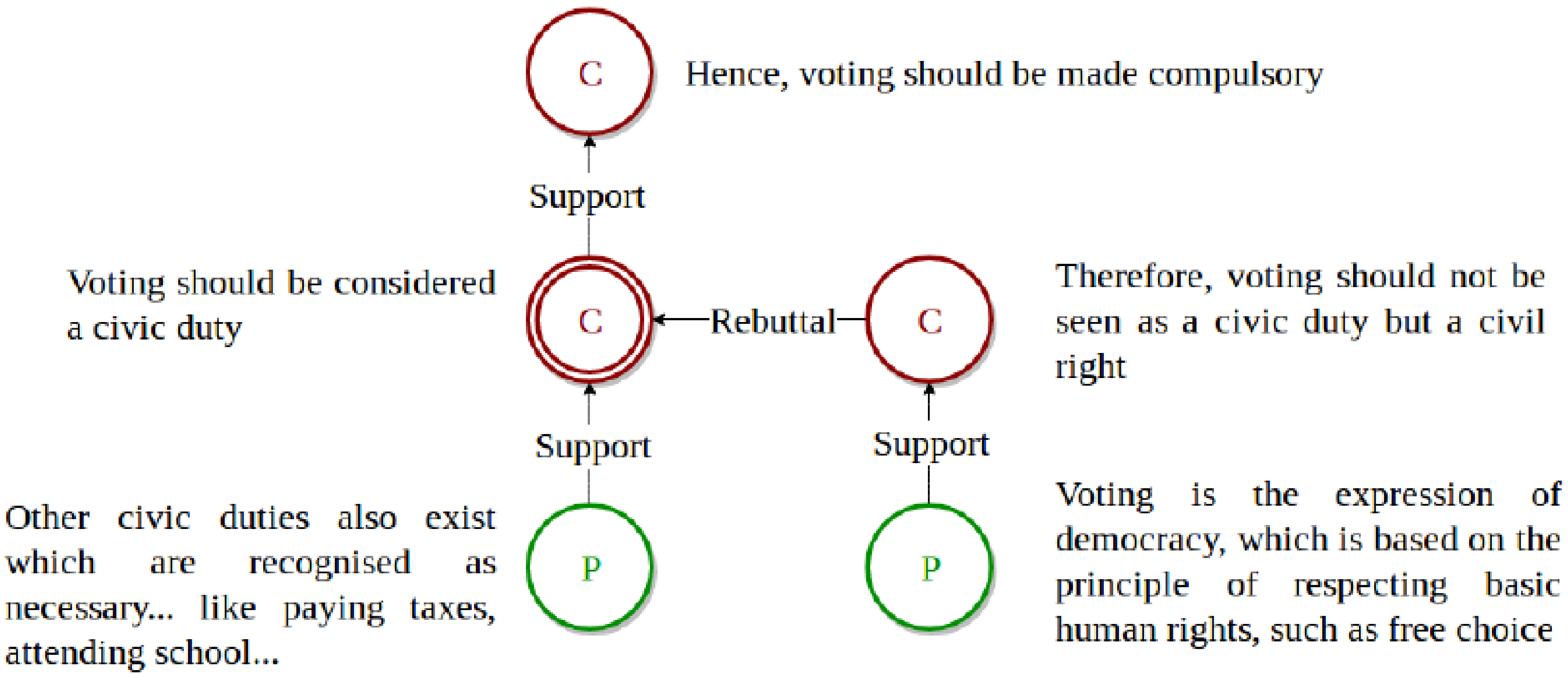
Niklaus et al., A Sentence Simplification System for Improving Relation Extraction, COLING (2017)

Argumentation Structures

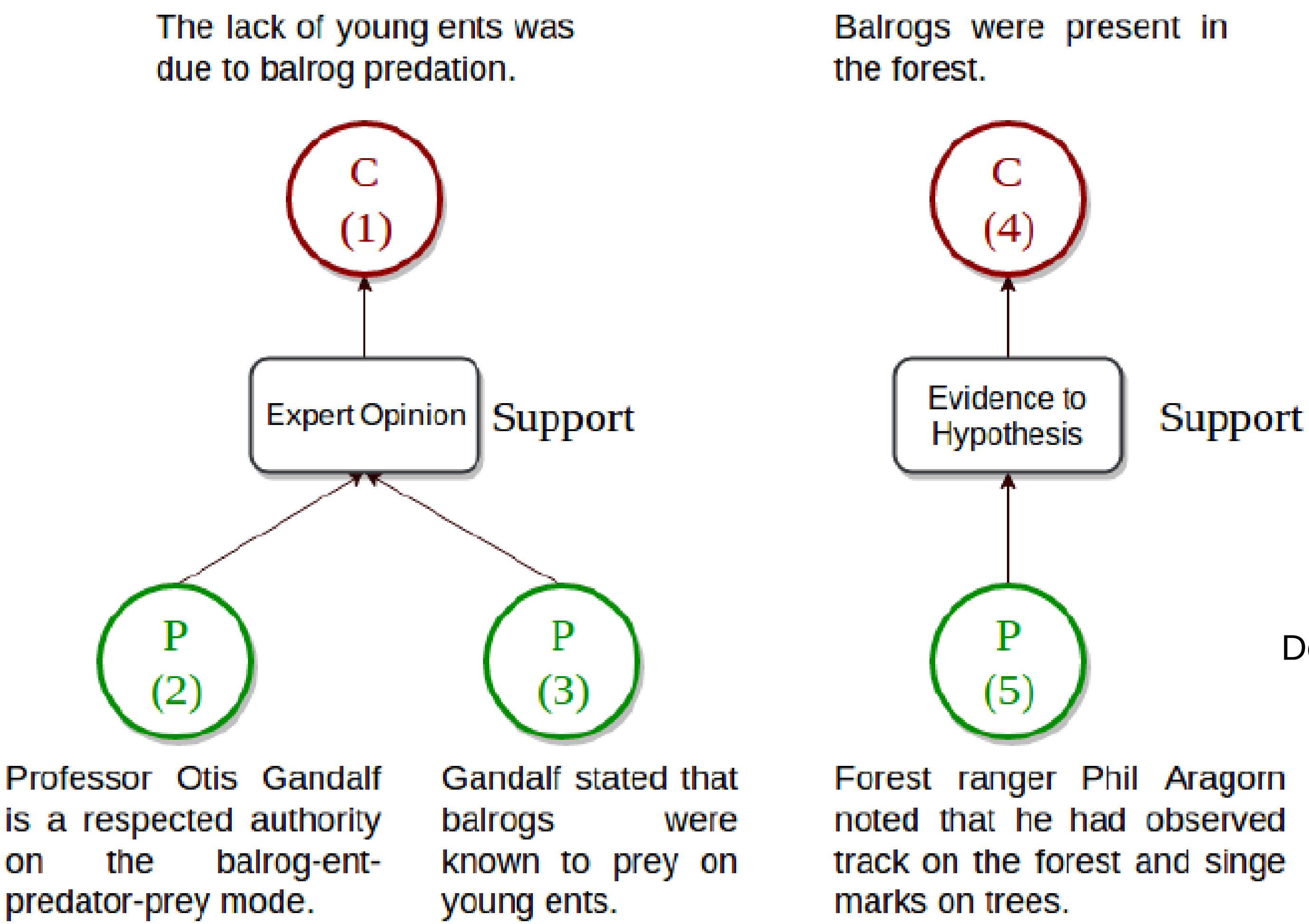


Stab & Gurevych, Parsing Argumentation Structures in Persuasive Essays, 2016.

Argumentative Discourse Unit Classification

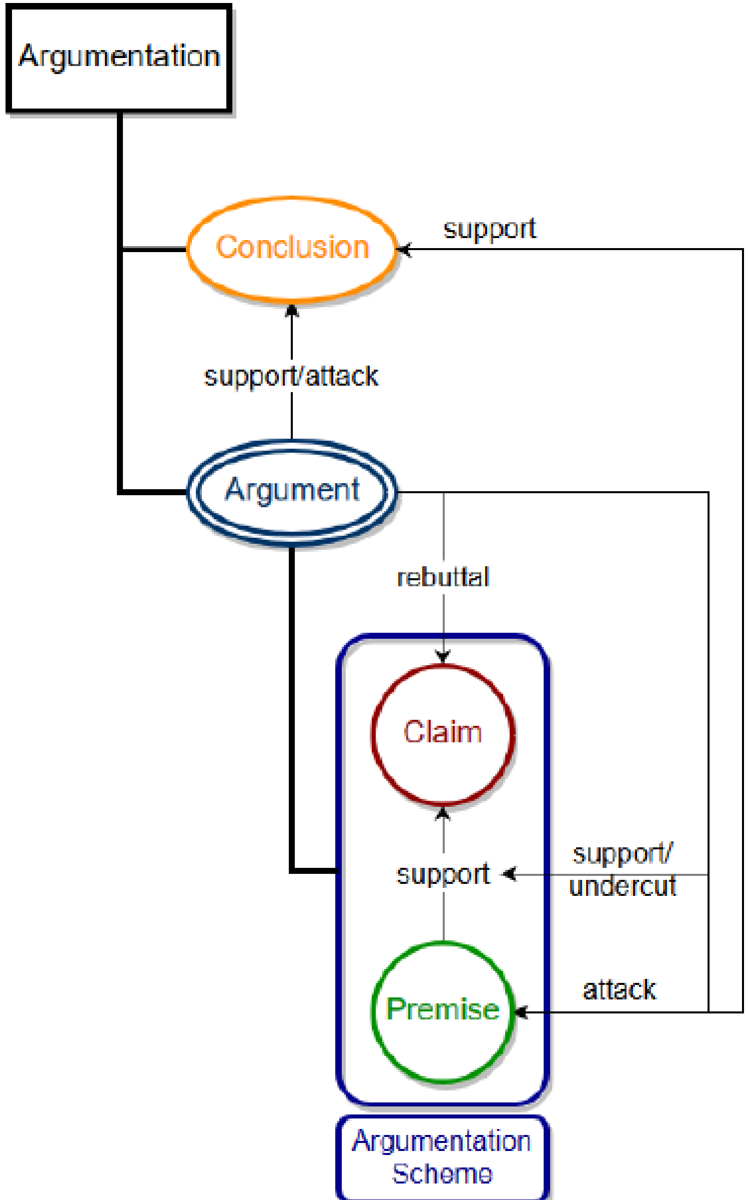


Argumentation Schemes



Douglas Walton

Unified Schema



Argument Mining Approaches

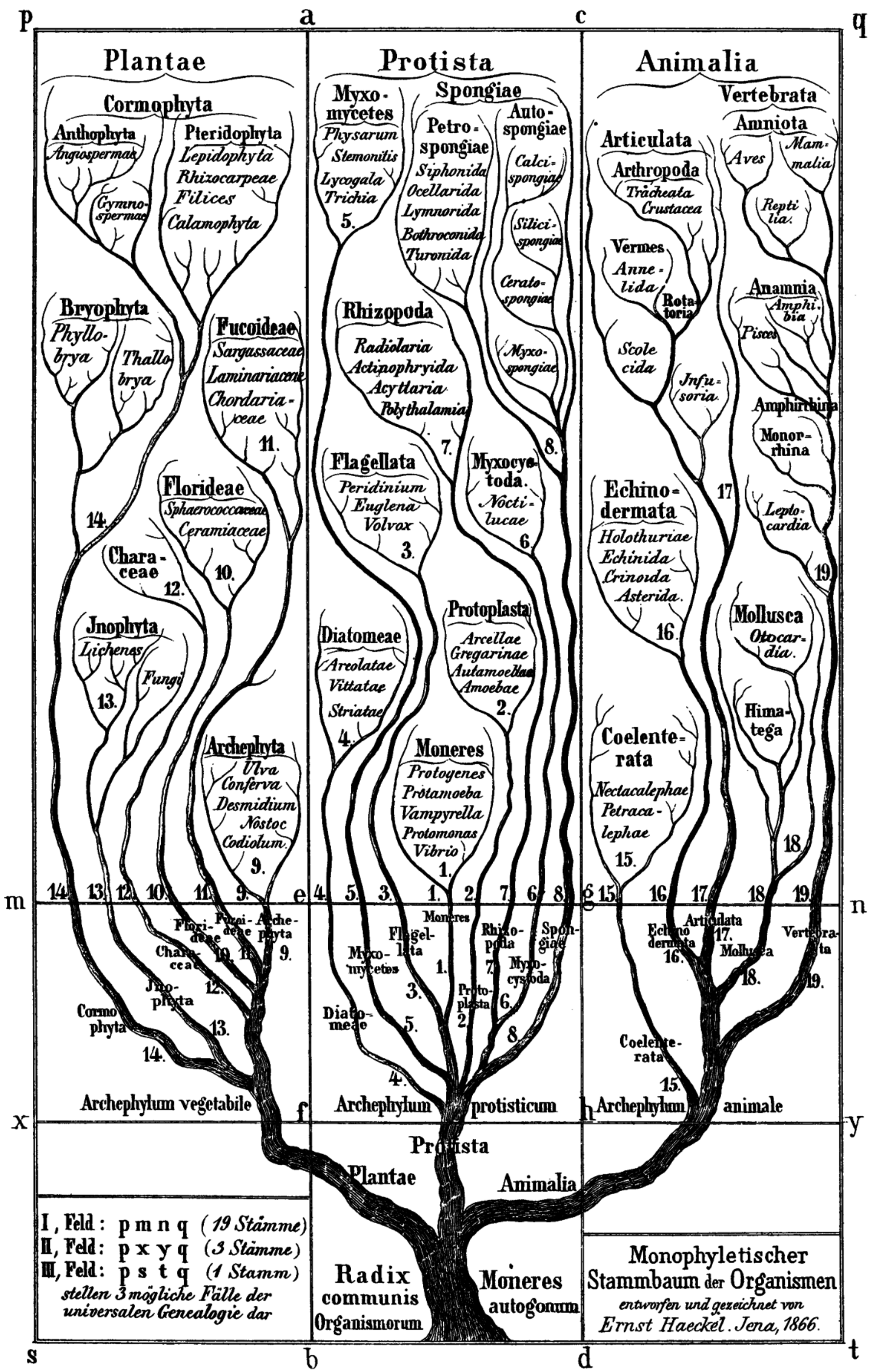
The University of Manchester

What to expect?

F1-score: 0.74

Stab & Gurevych, Parsing Argumentation Structures in Persuasive Essays, 2016.

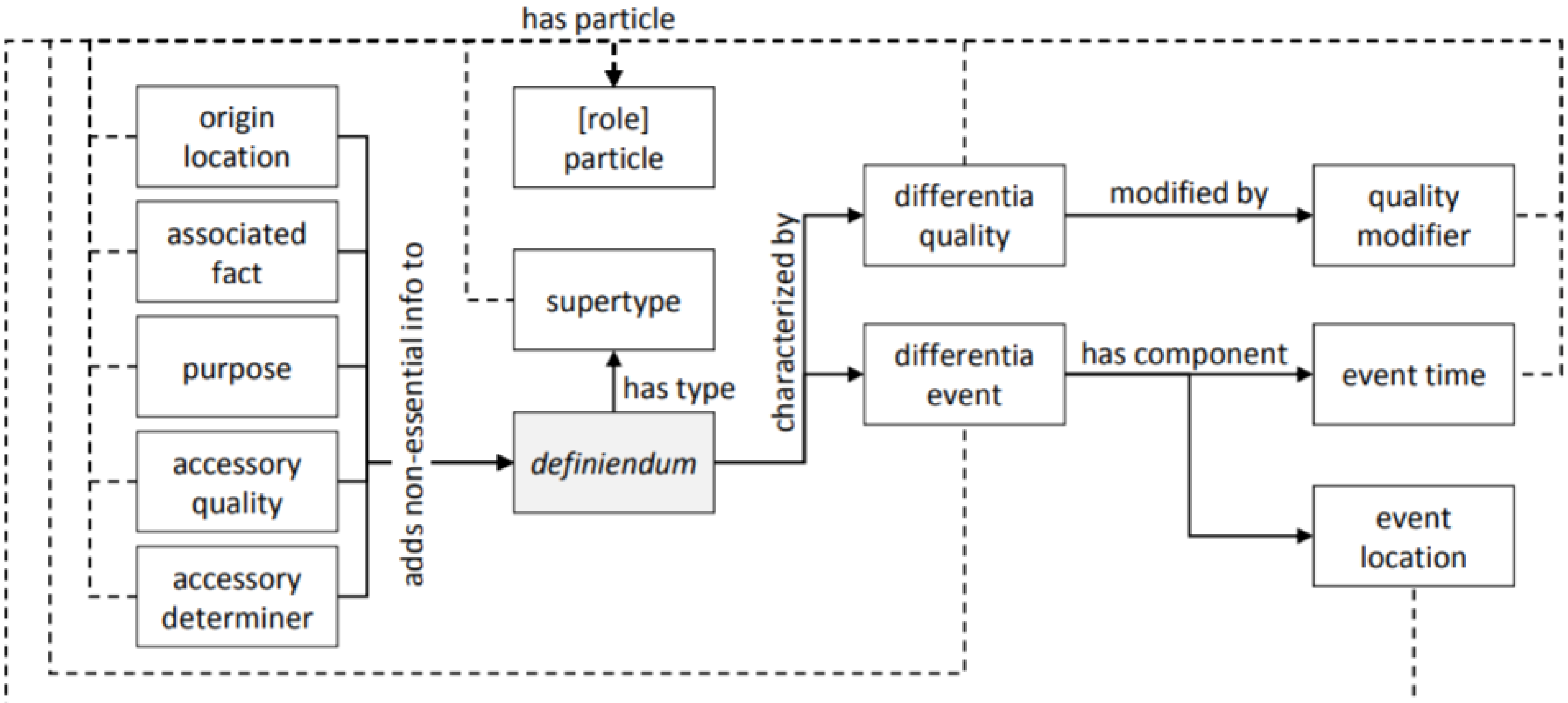
Task	AM System	Classifier	Corpus
1	Moens et al. 2007 [32]	Multinomial Naïve Bayes	AraucariaDB
1	Rooney et al. 2012 [40]	Convolution Kernel Methods	AraucariaDB
1	Florou et al. 2013 [13]	Decision Tree	Greek Corpus
1-2	Levy et al. 2014 [28]	Multi-Classifier Pipeline	Wikipedia
1-2	Lippi and Torroni 2015 [29]	SVM	IBM Corpus
1-2	Rinott et al. 2015 [39]	Multi-Classifier Pipeline	Wikipedia
2	Palau and Moens 2009 [34]	Max. Entropy and SVM	AraucariaDB & ECHR
2	Biran and Rambow 2011 [4]	Rule based + Naïve Bayes	RST Treebank, Wikipedia, blog threads from LiveJournal [1]
2	Habernal and Gurevych 2015 [20]	SVM-HMM	User Generated Content & Newspaper editorials
2-3	Stab and Gurevych 2014 [46]	SVM	Student Essays
1-2-3	Mochales and Moens 2011 [31]	Max. Entropy and SVM and CFG	AraucariaDB & ECHR
1-2-3	Stab and Gurevych 2016 [47]	Classifier Pipeline (CRF, SVM)	Student Essays
1-2-3	Daxenberger et al. 2017	RNN-LSTM-ER	Student Essays
4	Feng and Hirst 2011 [12]	C4.5 Decision Tree	AraucariaDB (subset)
4	Lawrence and Reed 2016 [26]	Naïve Bayes, SVM, Decision Tree	AraucariaDB (subset)



Definition-based Models

Semantic Roles for Lexical Definitions

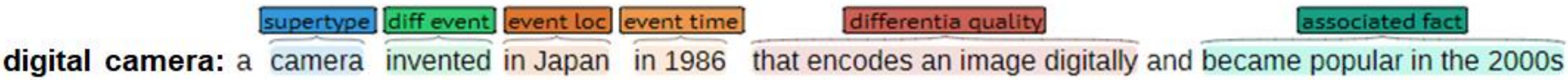
Aristotle's classic theory of definition introduced important aspects such as the **genus-differentia definition pattern** and the **essential/non-essential property differentiation**.

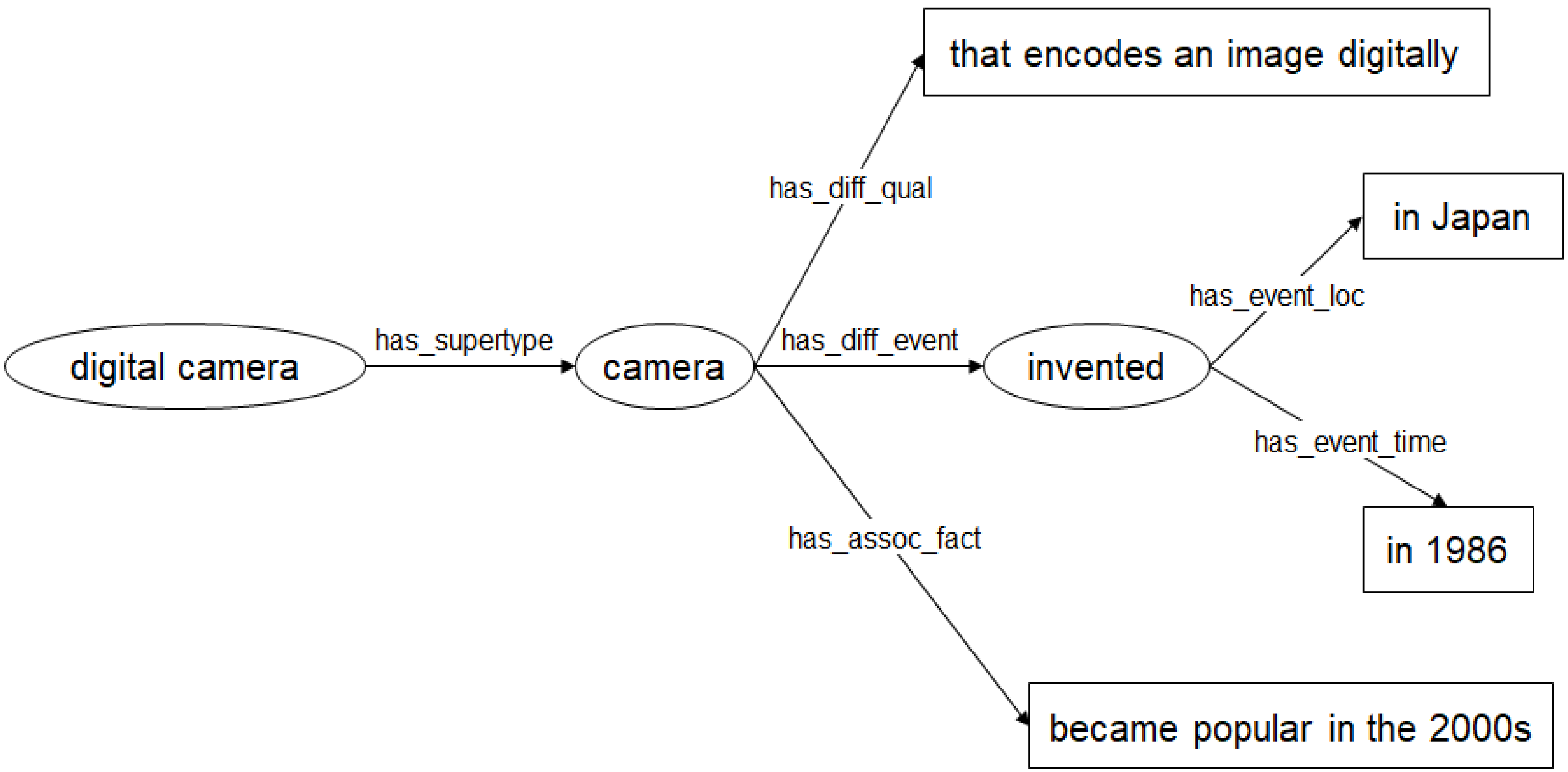


MANCHESTER 1824
The University of Manchester

Building the Definition Graph

Dictionary definitions are split into ***entity-centered semantic roles*** and converted to RDF





Data: WordNetGraph

RDF graph generated from WordNet.

https://github.com/Lambda-3/WordnetGraph

Silva et al., Categorization of Semantic Roles for Dictionary Definitions.
Cognitive Aspects of the Lexicon CogALex@COLING, 2017.

Emerging perspectives

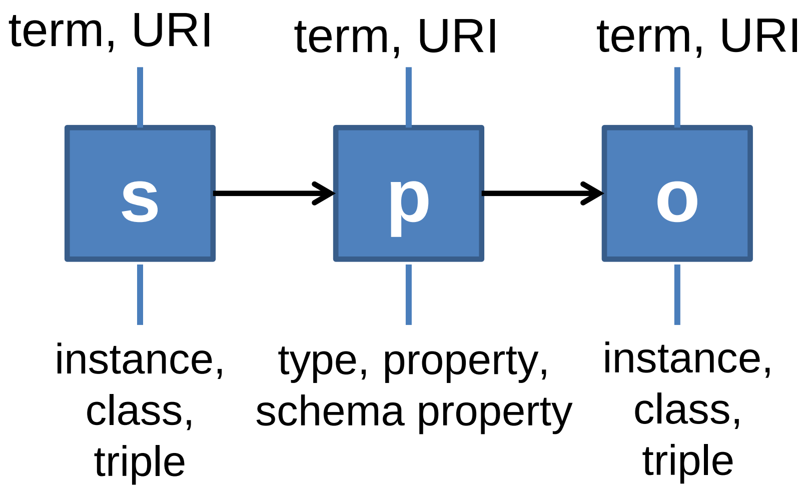
- The evolution of parsing and classification methods in NLP is inducing a new lightweight semantic representation.
- This representation dialogues with elements from logics, linguistics and the Semantic/Linked Data Web (especially RDF).
- However, they relax the semantic constraints of previous models (which were operating under assumptions for deductive reasoning or databases).

Emerging perspectives

- Knowledge graphs as lexical semantic models operating under a semantic best-effort mode (canonical identifiers when possible, otherwise, words).
- Possibly closer to the surface form of the text.
- Priority is on segmenting, categorizing and when possible, integrating.
- A representation (data model) convenient for AI engineering.

Categorization

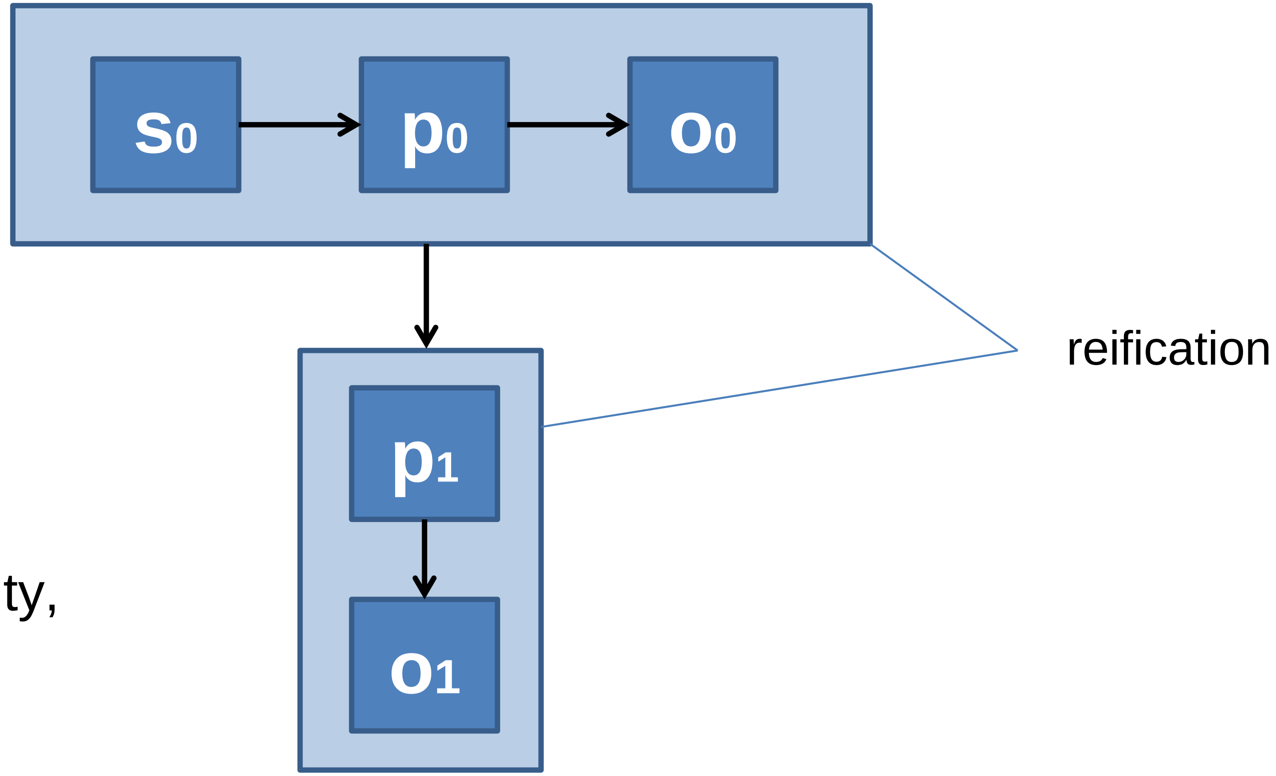
A fact (main clause):



* Can be a taxonomic fact.

Categorization

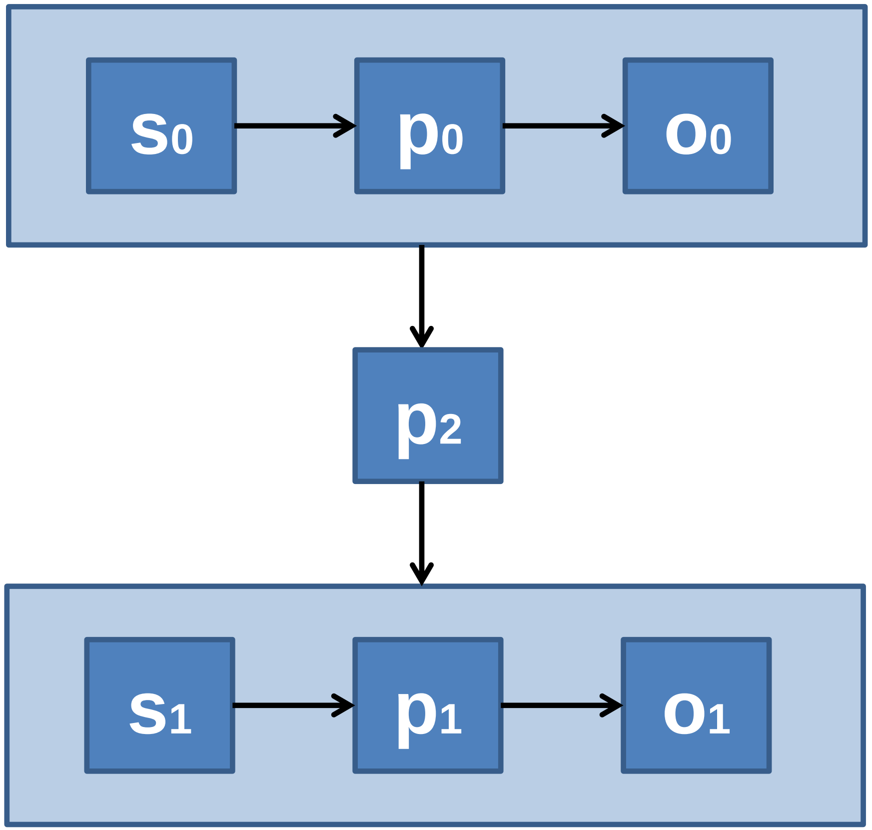
A fact with a context:



e.g.
- subordination (modality, temporality, spatiality, RSTs)
- fact probability
- polarity

Categorization

Coordinated facts:



e.g.

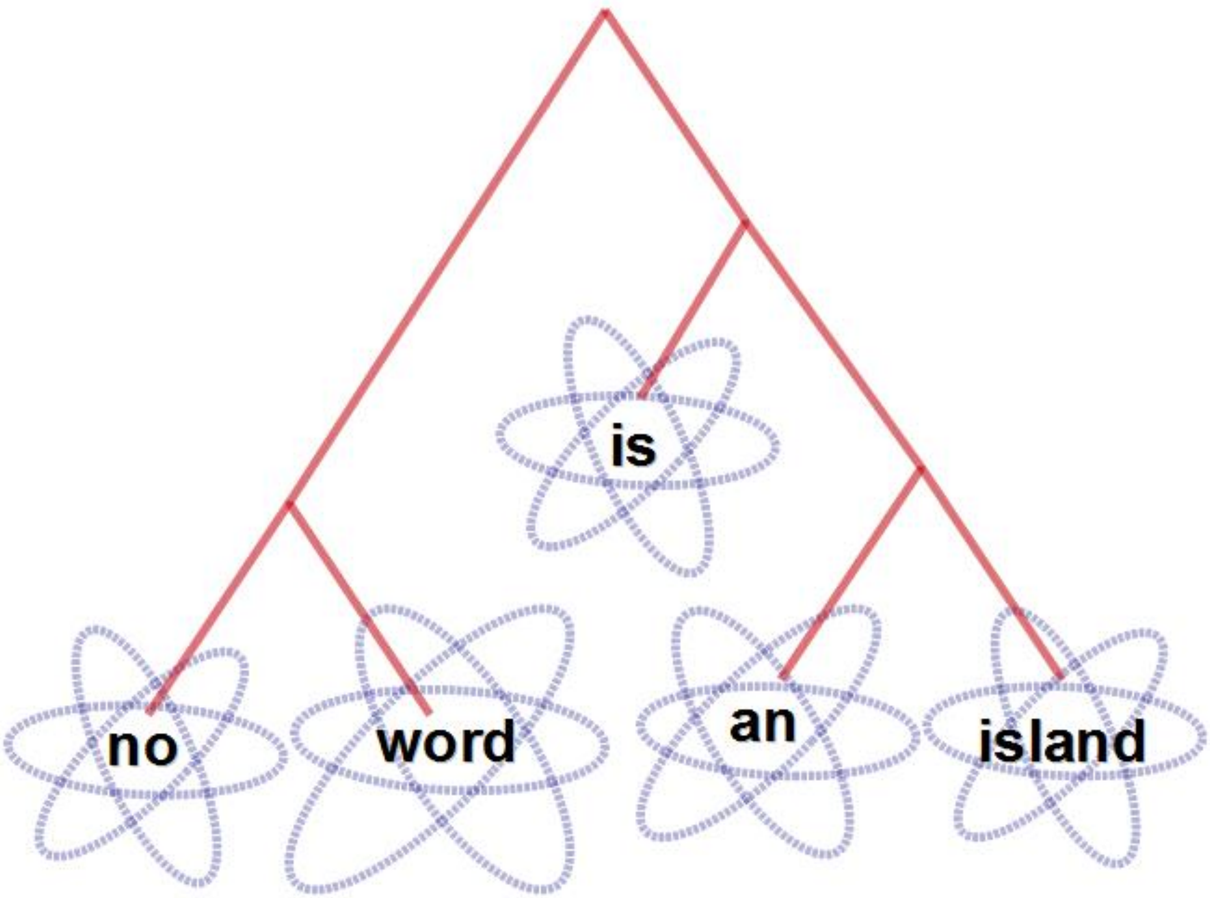
- coordination
- RSTs
- ADU

RDF-NL

https://github.com/Lambda-3/Graphene/blob/master/wiki/RDFNL-Format.md

Knowledge Graphs & Distributional Semantics

(A marriage made in heaven?)



Distributional Semantics

Distributional Semantic Models (Word Vector Models)

- Computational models that build contextual semantic representations from corpus data.

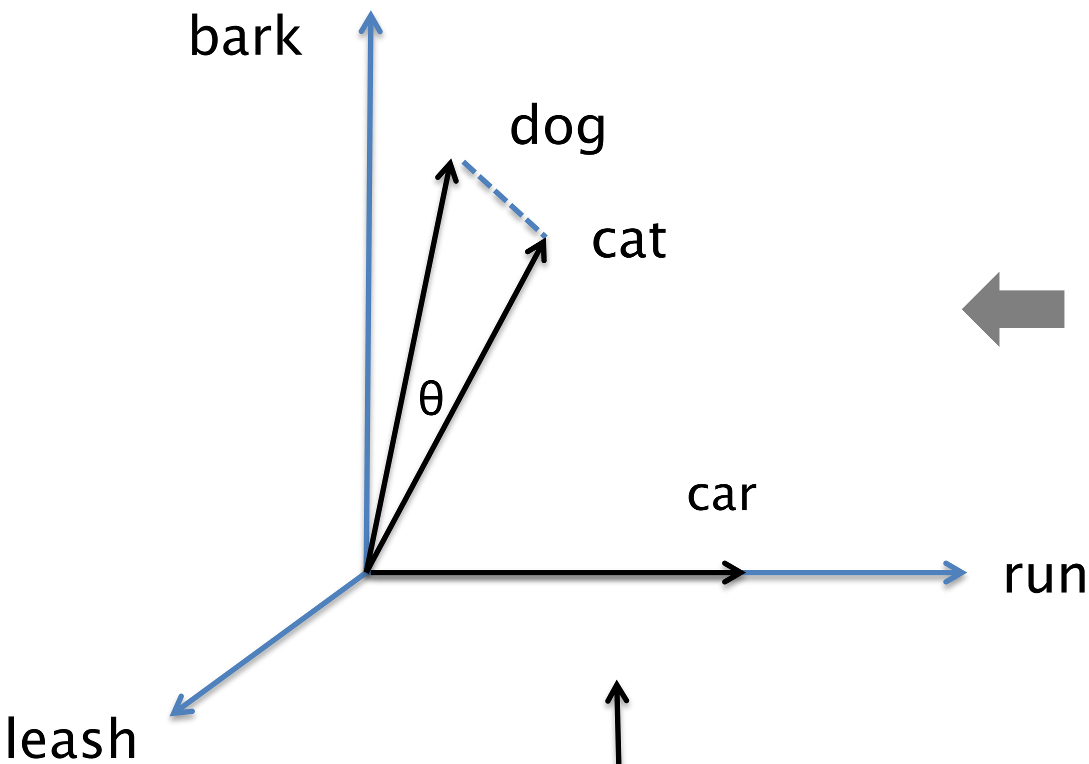
- **Semantic context** is represented by a **vector.**

- Vectors are obtained through the statistical analysis of the linguistic contexts of a word.

- Salience of contexts (cf. context weighting scheme).

- Semantic similarity/relatedness as the core operation over the model.

Distributional Semantics as Commonsense Knowledge



Commonsense is here

Semantic Approximation is here

Semantic Model with low acquisition effort

Measure	Definition
Euclidean	$\frac{1}{1+\sqrt{\sum_{i=1}^{n}(u_i-v_i)^2}}$
Cityblock	$\frac{1}{1+\sum_{i=1}^{n}\lvert u_i-v_i\rvert}$
Chebyshev	$\frac{1}{1+\max_i \lvert u_i-v_i\rvert}$
Cosine	$\frac{u\cdot v}{\lvert u\rvert \lvert v\rvert}$
Correlation	$\frac{(u-\mu_u)\cdot(v-\mu_v)}{\lvert u\rvert \lvert v\rvert}$
Dice	$\frac{2\sum_{i=0}^{n} min(u_i,v_i)}{\sum_{i=0}^{n} u_i+v_i}$
Jaccard	$\frac{u\cdot v}{\sum_{i=0}^{n} u_i+v_i}$
Jaccard2	$\frac{\sum_{i=0}^{n} min(u_i,v_i)}{\sum_{i=0}^{n} max(u_i,v_i)}$
Lin	$\frac{\sum_{i=0}^{n} u_i+v_i}{\lvert u\rvert+\lvert v\rvert}$
Tanimoto	$\frac{u\cdot v}{\lvert u\rvert+\lvert v\rvert-u\cdot v}$
Jensen-Shannon Div	$1-\frac{\frac{1}{2}(D(u\Vert\frac{u+v}{2})+D(v\Vert\frac{u+v}{2}))}{\sqrt{2\log 2}}$
α-skew	$1-\frac{D(u\Vert\alpha v+(1-\alpha)u)}{\sqrt{2\log 2}}$

Similarity Measures

X

Scheme	Definition
None	$w_{ij}=f_{ij}$
TF-IDF	$w_{ij}=\log(f_{ij})\times\log(\frac{N}{n_j})$
TF-ICF	$w_{ij}=\log(f_{ij})\times\log(\frac{N}{f_j})$
Okapi BM25	$w_{ij}=\frac{f_{ij}}{0.5+1.5\times\frac{f_j}{\frac{f_j}{j}}+f_{ij}}\log\frac{N-n_j+0.5}{f_{ij}+0.5}$
ATC	$w_{ij}=\frac{(0.5+0.5\times\frac{f_{ij}}{max_f})\log(\frac{N}{n_j})}{\sqrt{\sum_{i=1}^{N}[(0.5+0.5\times\frac{f_{ij}}{max_f})\log(\frac{N}{n_j})]^2}}$
LTU	$w_{ij}=\frac{(\log(f_{ij})+1.0)\log(\frac{N}{n_j})}{0.8+0.2\times f_j\times\frac{j}{f_j}}$
MI	$w_{ij}=\log\frac{P(t_{ij}\mid c_j)}{P(t_{ij})P(c_j)}$
PosMI	$\max(0,\mathrm{MI})$
T-Test	$w_{ij}=\frac{P(t_{ij}\mid c_j)-P(t_{ij})P(c_j)}{\sqrt{P(t_{ij})P(c_j)}}$
χ^2	see (Curran, 2004, p. 83)
Lin98a	$w_{ij}=\frac{f_{ij}\times f}{f_i\times f_j}$
Lin98b	$w_{ij}=-1\times\log\frac{n_j}{N}$
Gref94	$w_{ij}=\frac{\log f_{ij}+1}{\log n_j+1}$

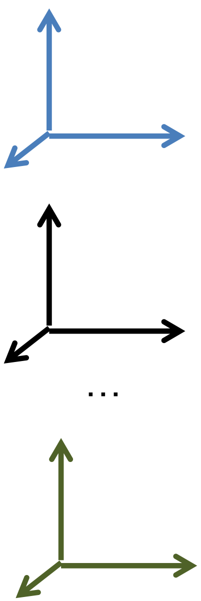
Context Weighting Measures

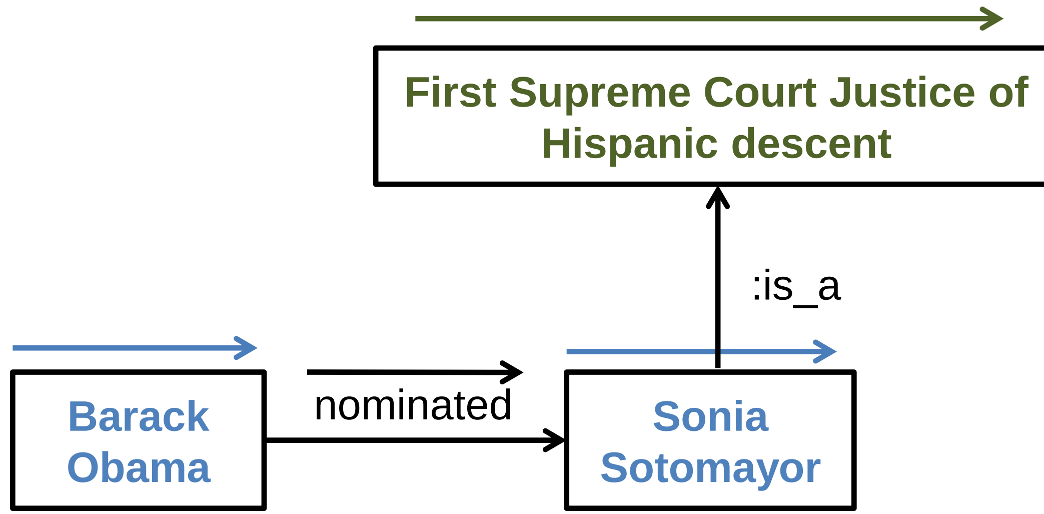
Kiela & Clark, 2014

… and of course, Glove and W2V

Distributional-Relational Models

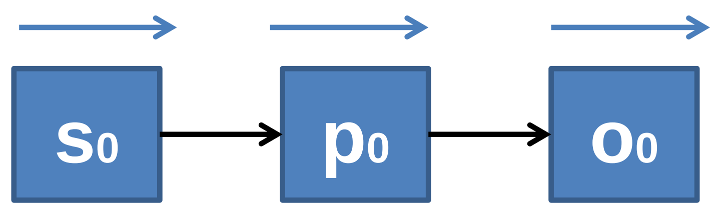
LSA, ESA, W2V, GLOVE, …



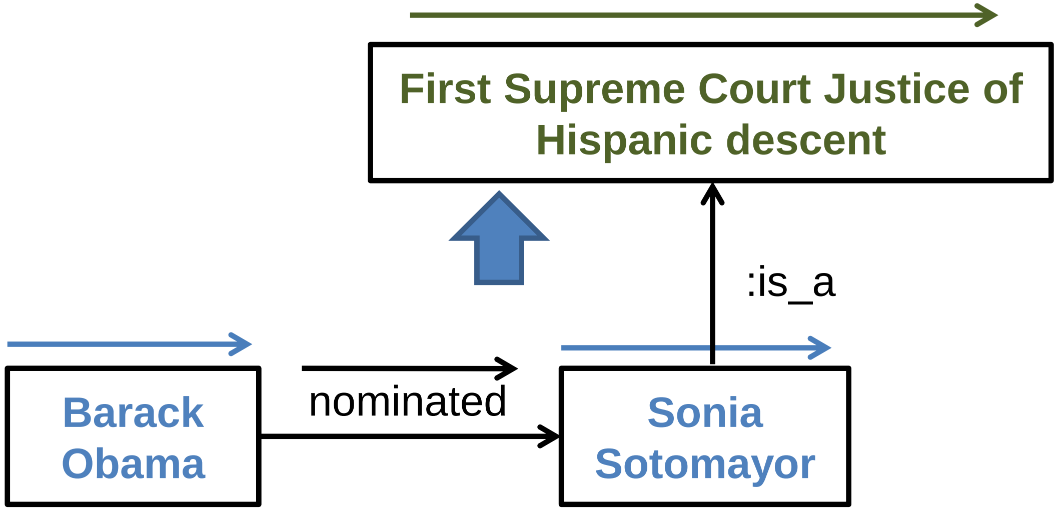


Distributional Relational Networks, AAAI Symposium (2013).

A Compositional-Distributional Semantic Model for Searching Complex Entity Categories, ACL *SEM (2016)

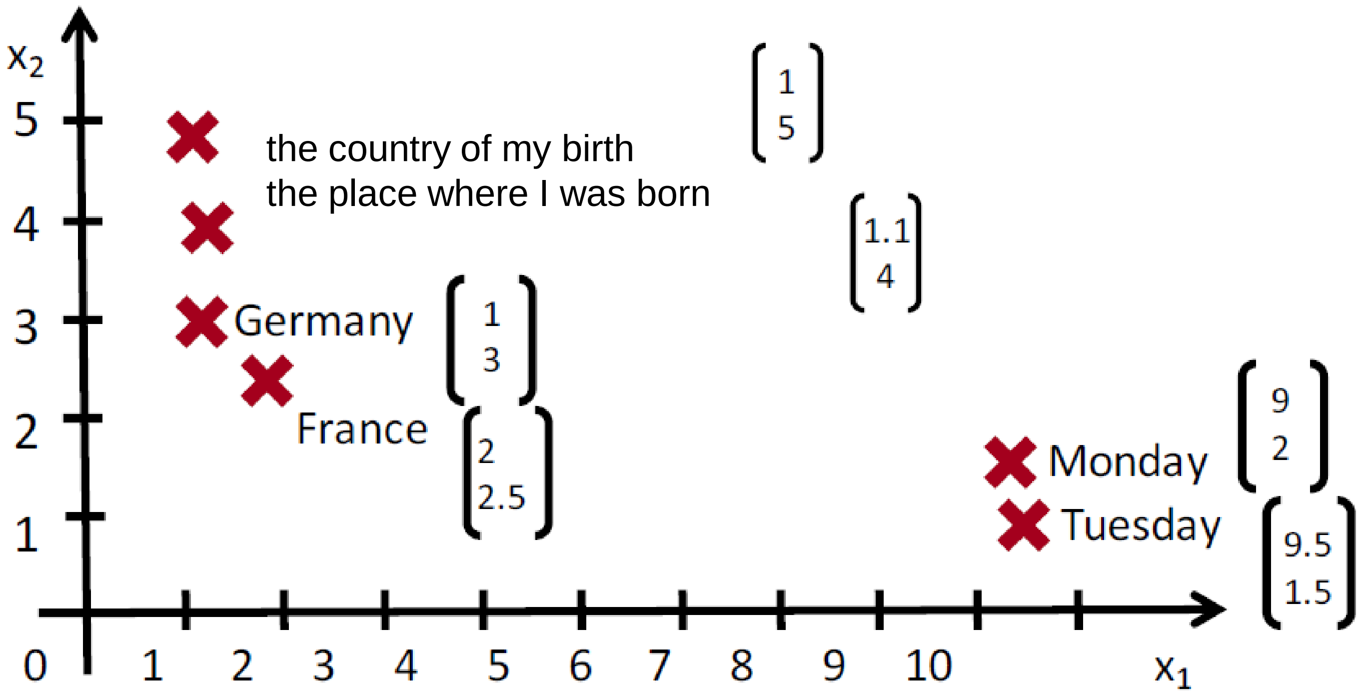


Compositionality of Complex Nominals



Original	*Paraphrased*
Prehistoric Canines	Ancestral Wolves
Soviet Pop Music Groups	Popular Musical Bands in the USSR
American Architectural Styles	Fashions of American Building Design
Defunct Companies of Finland	Bankrupt Finnish Businesses

Building on Word Vector Space Models

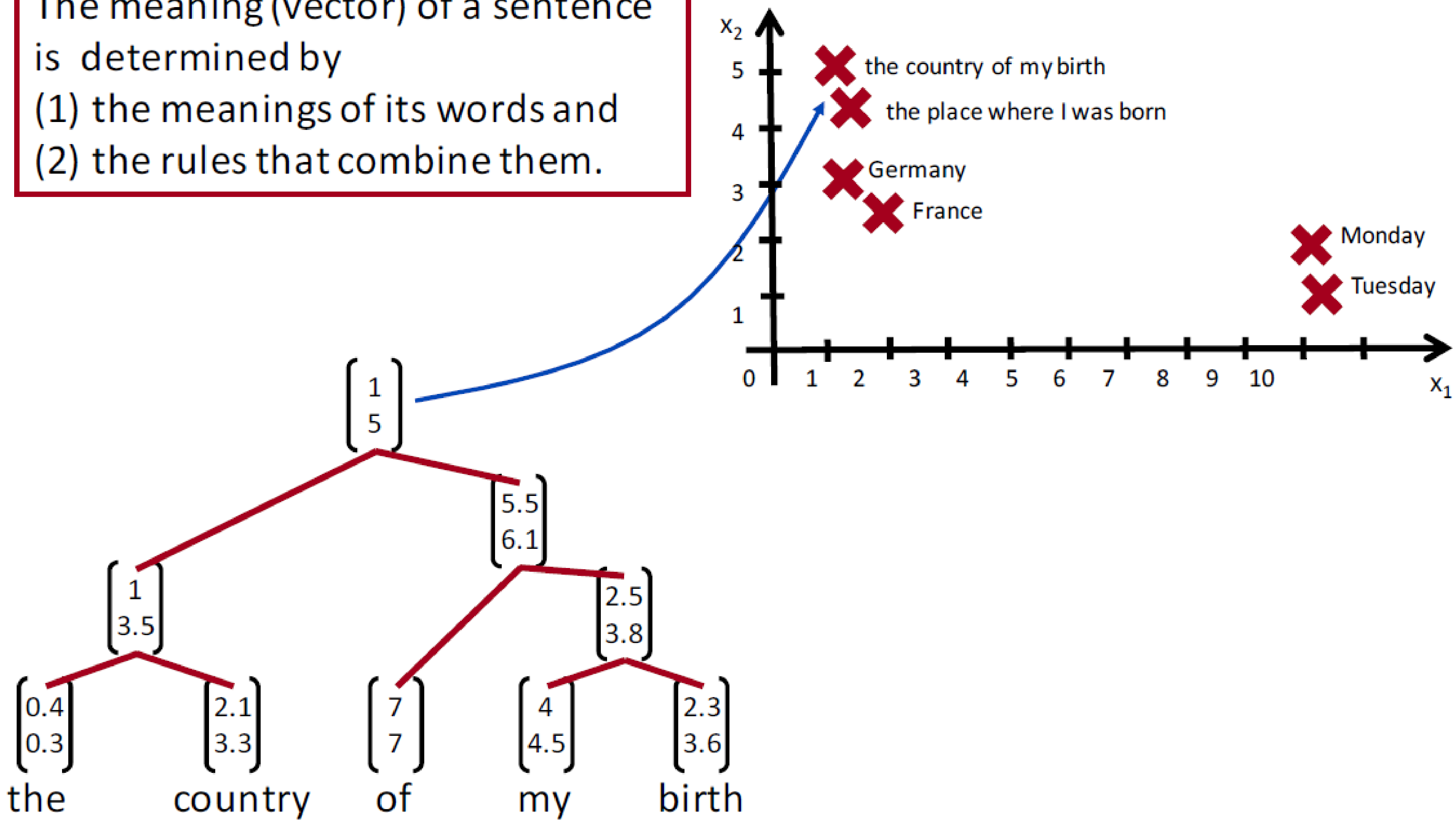


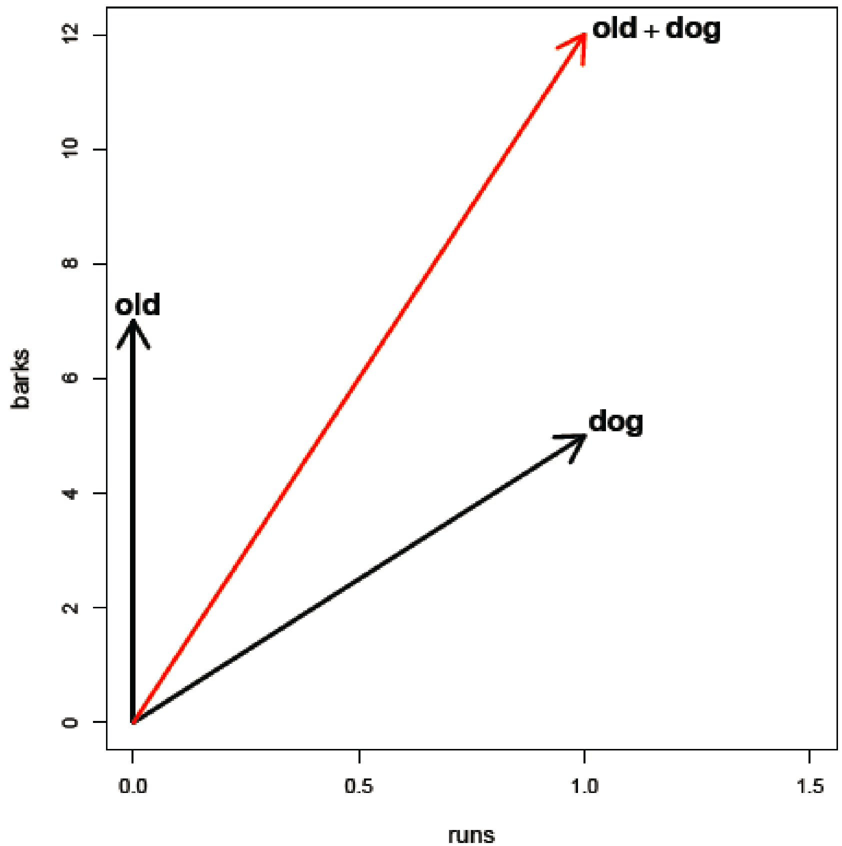
- But how can we represent the meaning of longer phrases?
- By mapping them into the same vector space!

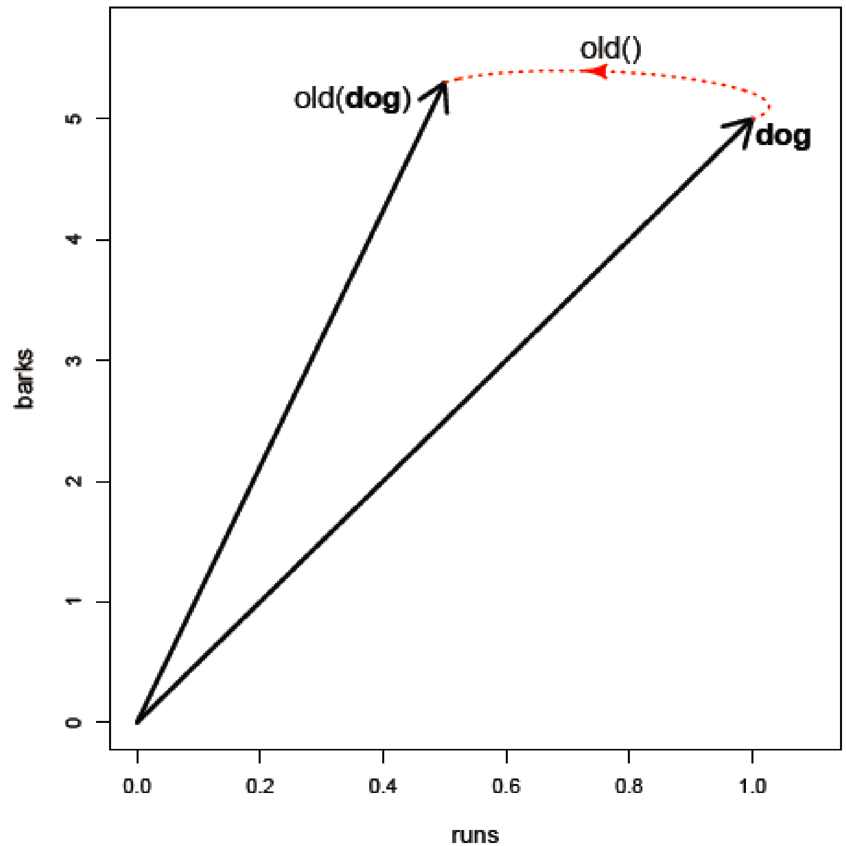
Recursive Neural Networks

Use principle of compositionality

The meaning (vector) of a sentence is determined by
(1) the meanings of its words and
(2) the rules that combine them.

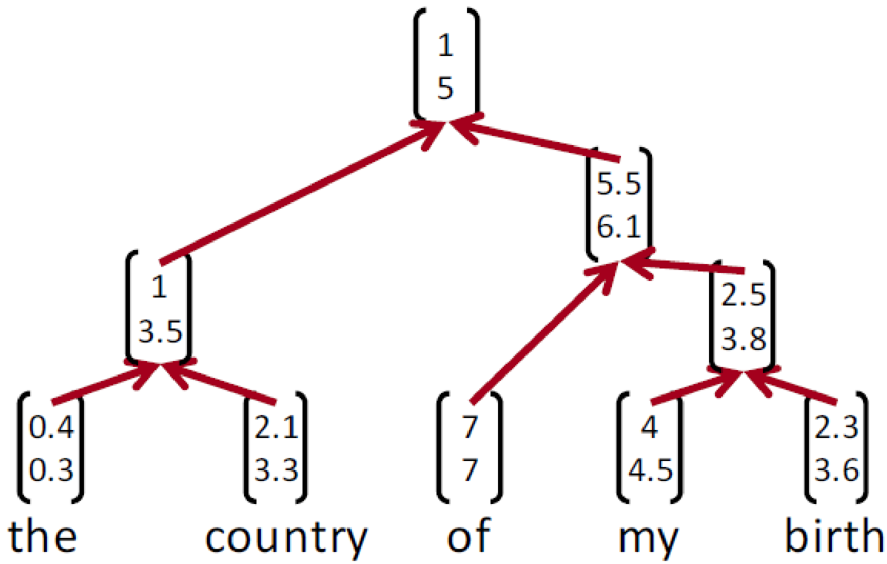


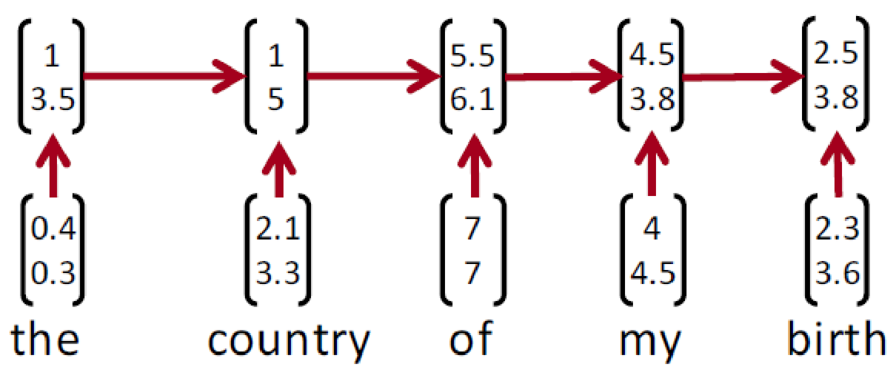




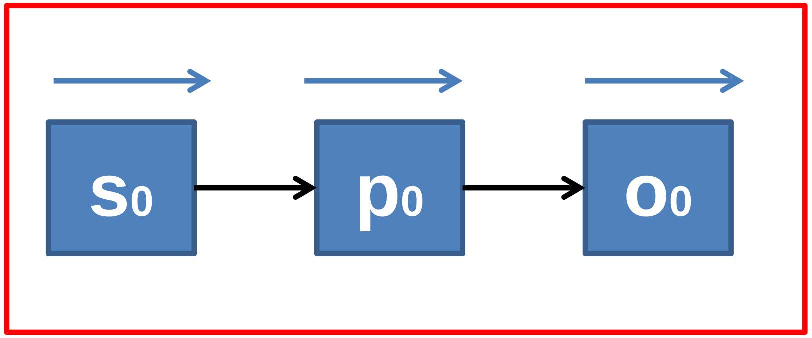
A Compositional-Distributional Semantic Model for Searching Complex Entity Categories, *SEM (2016)

Recursive vs recurrent neural networks

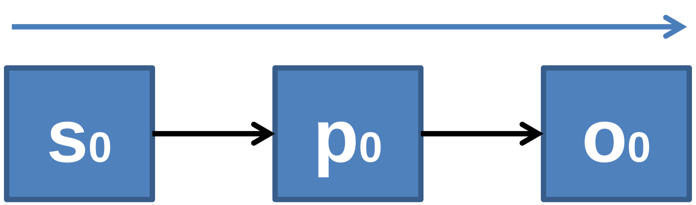




Segmented Spaces vs Unified Space



- Assumes is <s,p,o> naturally irreconcilable.
- Inherent dimensional reduction mechanism.
- Facilitates the specialization of embedding-based approximations.



- Easier to compute identity.
- Requires complex and high-dimensional tensorial model.

Software: Indra

- Semantic approximation server
- Multi-lingual (12 languages)
- Multi-domain
- Different compositional models

https://github.com/Lambda-3/indra

Semantic Relatedness for All (Languages): A Comparative Analysis of Multilingual Semantic Relatedness using Machine Translation, EKAW, (2016).

The University of Manchester

"On our best behaviour"

Levesque, 2013

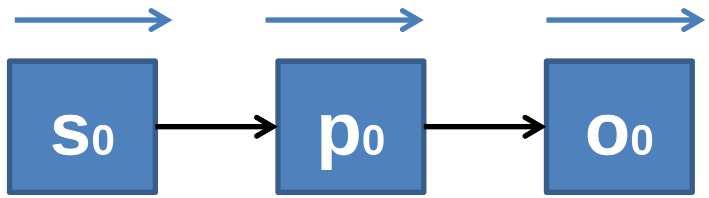
"It is not enough to build knowledge bases without paying closer attention to the demands arising from their use."

"We should explore more thoroughly **the space of computations between fact retrieval and full automated logical reasoning**."

How to access Distributional-Knowledge Graphs efficiently?

- Depends on the target operations in the Knowledge Graphs (more on this later).

How to access Distributional-Knowledge Graphs efficiently?

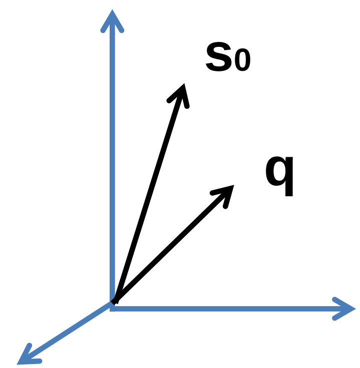


Database + IR

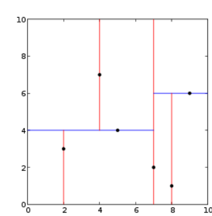
Structured Queries

Query planning
Cardinality
Indexing
Skyline
Bitmap indexes
…

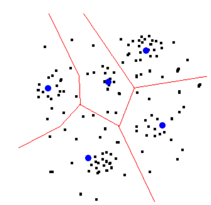
Approximation Queries



Inverted index
sharding
disk access
optimization
…



Multiple Randomized K-d Tree Algorithm



The Priority Search K-Means Tree algorithm

How to access Distributional-Knowledge Graphs efficiently?

$$s_0 \rightarrow p_0 \rightarrow o_0$$

Database + IR

Structured Queries

Approximation Queries

Software: StarGraph

- Distributional Knowledge Graph Database.
- Word embedding Database.

https://github.com/Lambda-3/Stargraph

Freitas et al., Natural Language Queries over Heterogeneous Linked Data Graphs: A Distributional-Compositional Semantics Approach, 2014.

Emerging perspectives

- Graph-based data models + Distributional Semantic Models (Word embeddings) have complementary semantic value.

- Graph-based Data Models:
  - Facilitates querying, integration and rule-based reasoning.

- Distributional Semantic Models:
  - Supports semantic approximation, coping with vocabulary variation.

Emerging perspectives

- AI systems require access to comprehensive background knowledge for semantic interpretation tasks.

- Inheriting from Information Retrieval and Databases:
  - General Indexing schemes,
  - Particular Indexing schemes,
    - Spatial, temporal, topological, probabilistic, causal, …
  - Query planning,
  - Data compression,
  - Distribution,
  - … even supporting hardware strategies.

Emerging perspectives

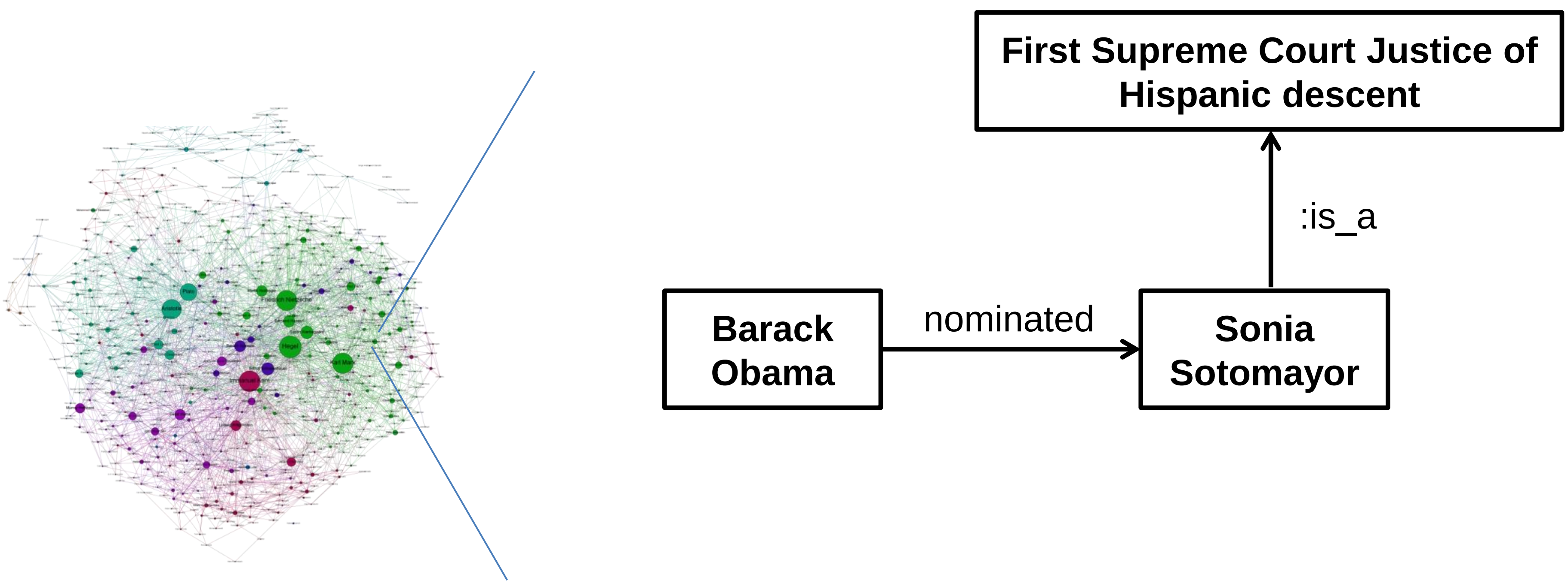
- **One size of embedding does not fit all**: Operate with multiple distributional + compositional models for different data model types (I, C, P), different domains and different languages.

MANCHESTER 1824

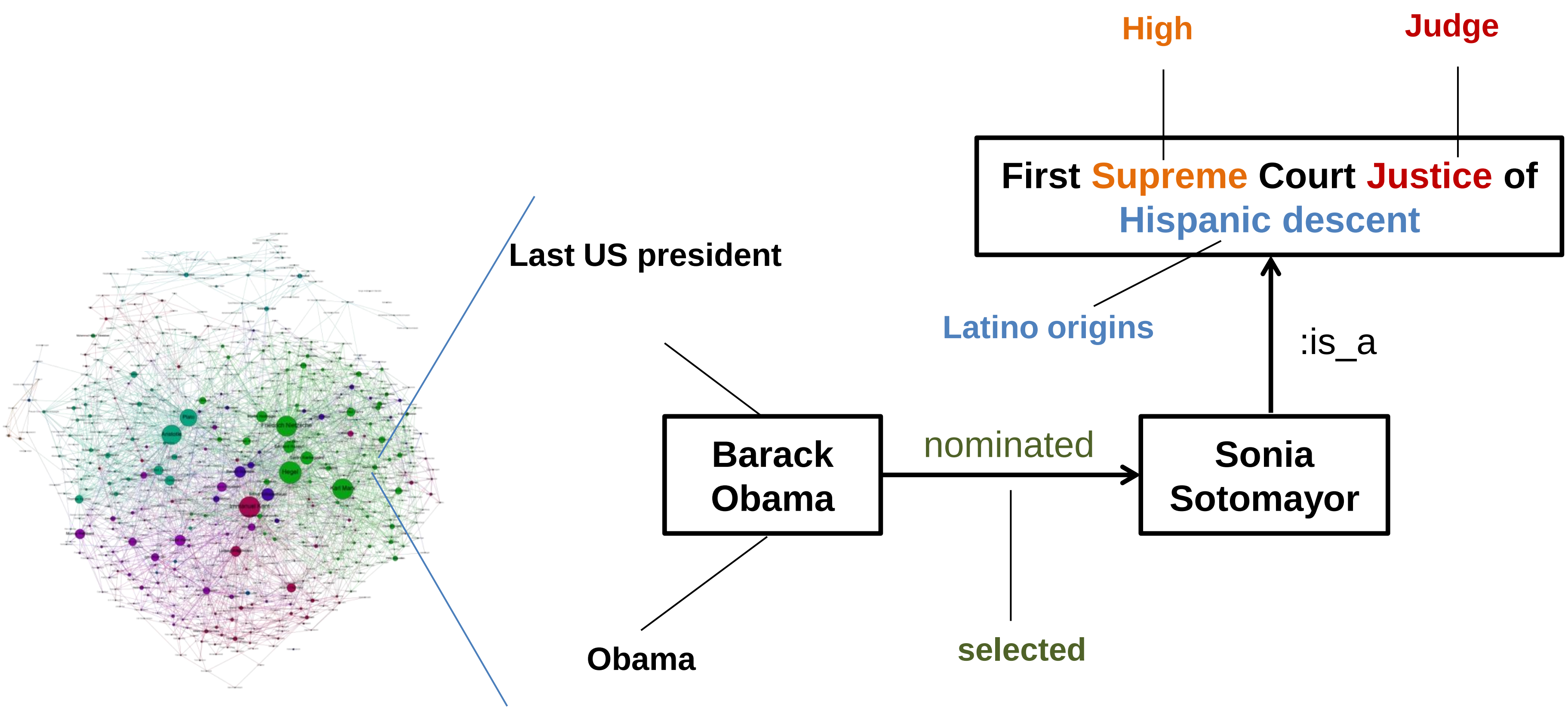
The University of Manchester

Effective Semantic Parsing for Large KBs

The Vocabulary Problem



The Vocabulary Problem



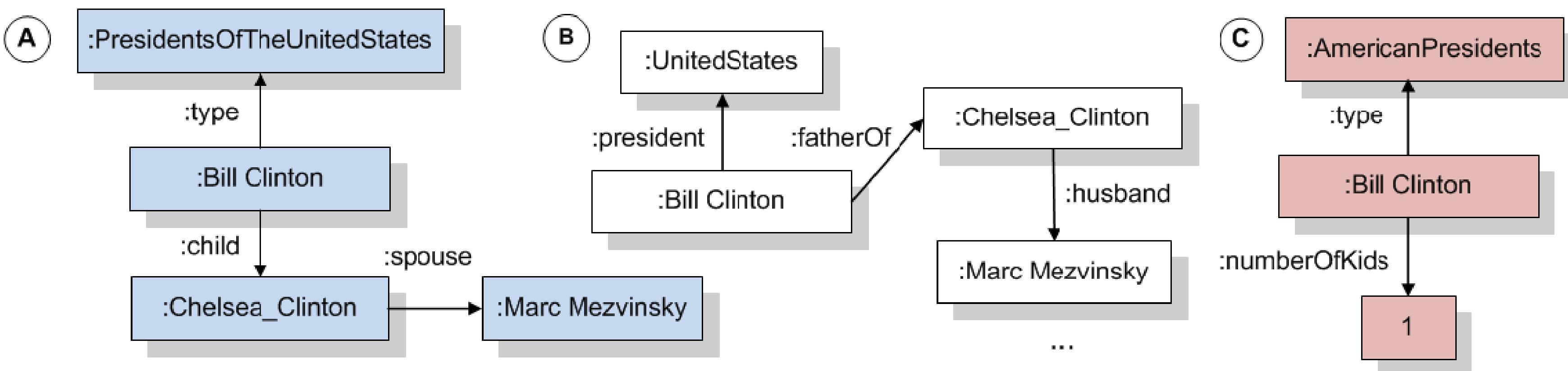
Vocabulary Problem for KGs

Query: Who is the daughter of Bill Clinton married to?

Semantic Gap

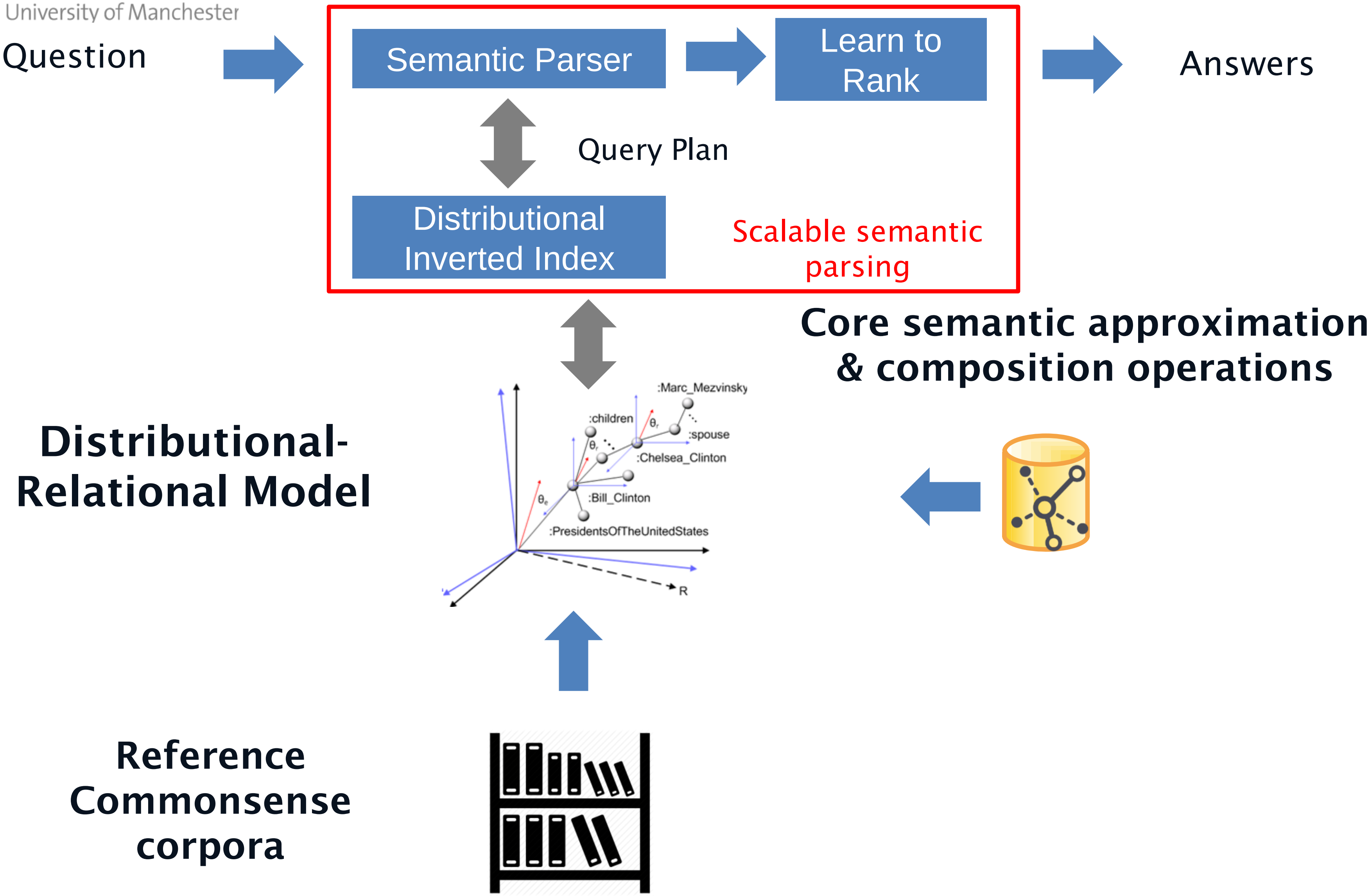
Schema-agnostic query mechanisms

Possible representations



- Abstraction level differences
- Lexical variation

- Structural (compositional) differences


Question
Semantic Parser
Learn to Rank
Answers
Query Plan
Distributional Inverted Index
Scalable semantic parsing
Core semantic approximation & composition operations
Distributional-Relational Model
:Marc_Mezvinsky
:children
:spouse
:Chelsea_Clinton
:Bill_Clinton
:PresidentsOfTheUnitedStates
R
Reference Commonsense corpora

Minimizing the Semantic Entropy for the Semantic Matching

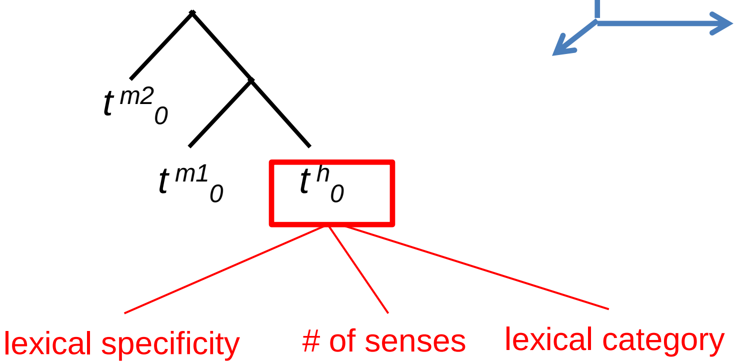
Definition of a semantic pivot: first query term to be resolved in the database.

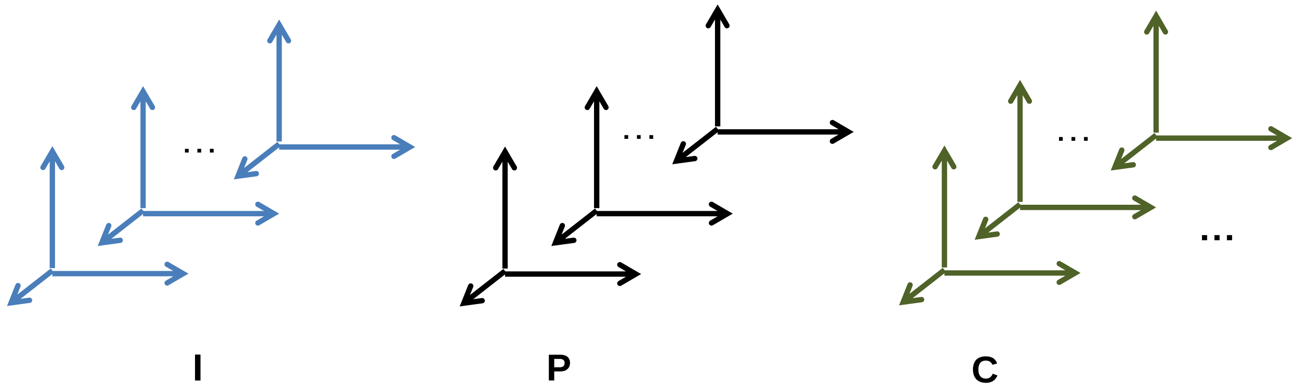
- Maximizes the reduction of the semantic configuration space.
- Less prone to more complex synonymic expressions and abstraction-level differences.
- Semantic pivot serves as interpretation context for the remaining alignments.
- proper nouns >> nouns >> complex nominals >> adjectives , verbs.

MANCHESTER
1824
The University of Manchester

$$\Gamma = \{I, P, C, V\}$$

$$\vec{q} = \vec{t^{\Gamma}}_0, \ldots, \vec{t^{\Gamma}}_n$$

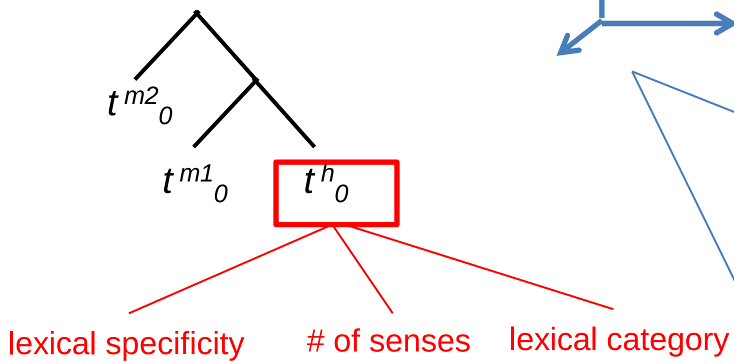


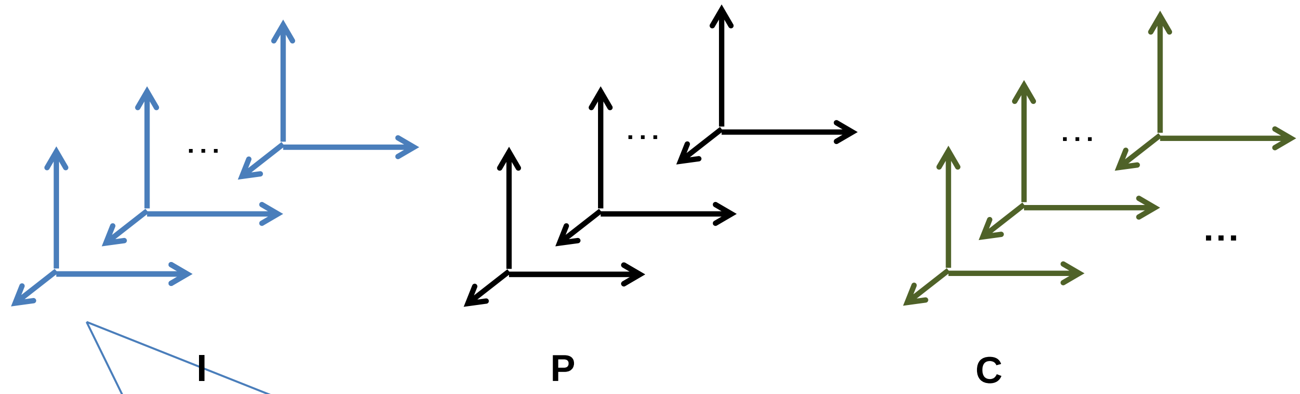


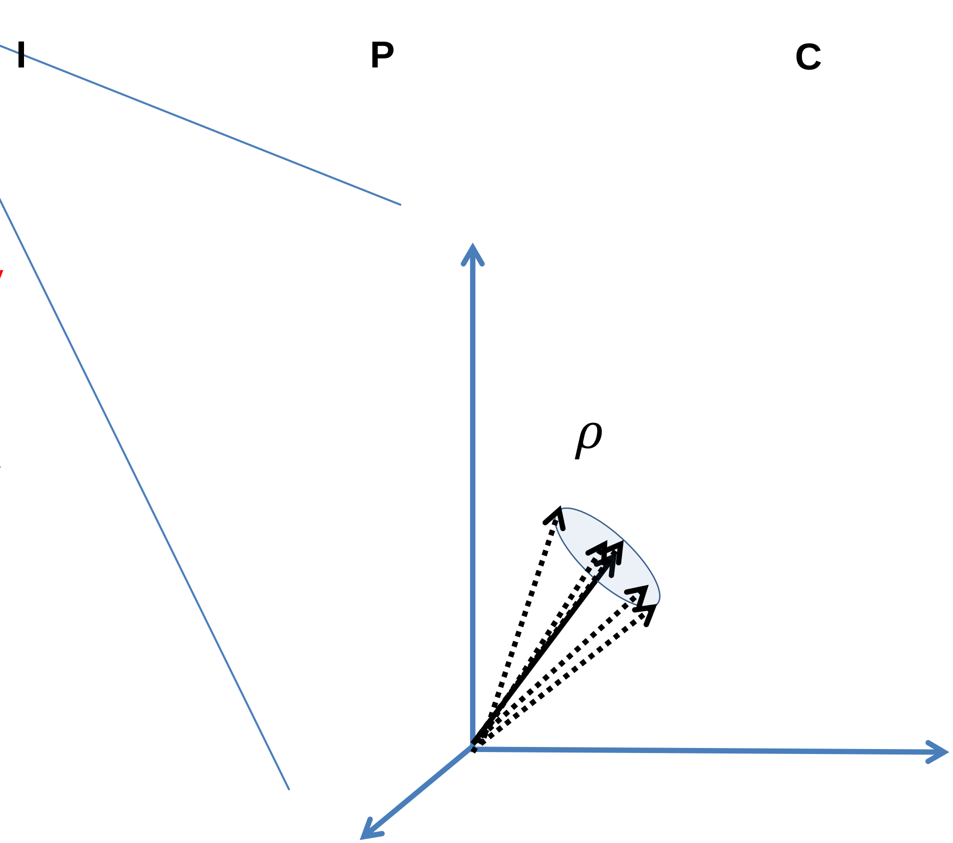
The University of Manchester

$$\Gamma = \{I, P, C, V\}$$

$$\vec{q} = \vec{t^{\Gamma}}_{0}, \ldots, \vec{t^{\Gamma}}_{n}$$





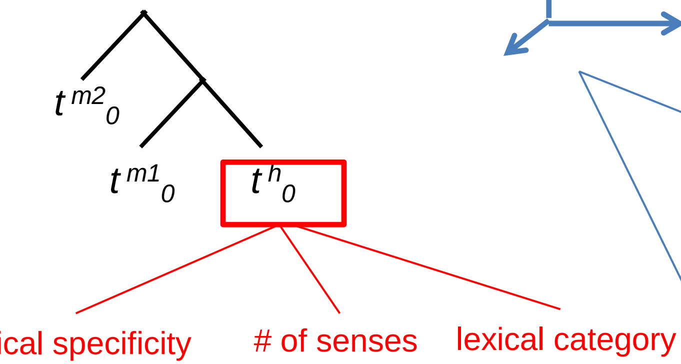


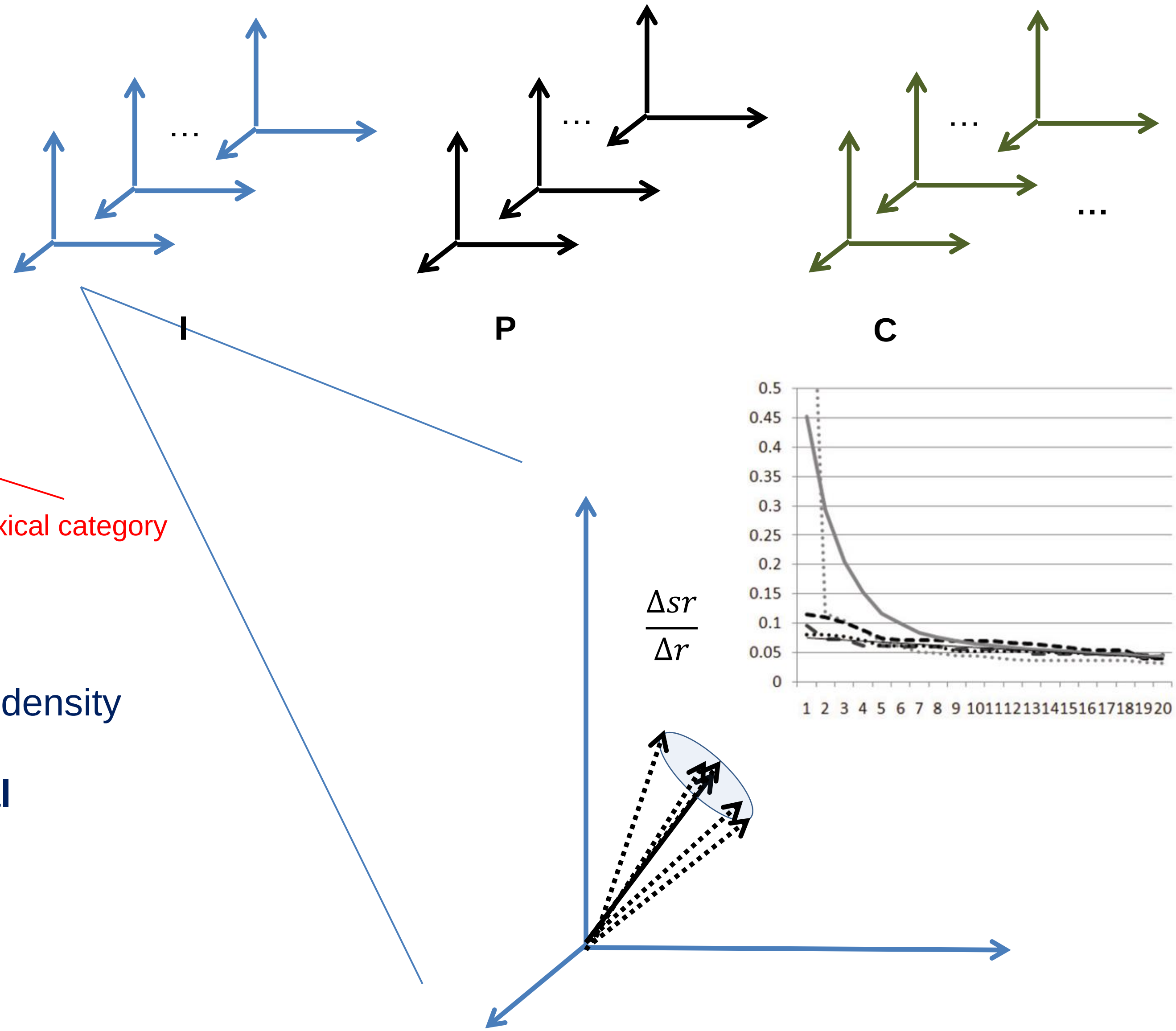
- **Vector neighborhood density**
- Semantic differential

MANCHESTER
1824
The University of Manchester

$$\Gamma = \{I, P, C, V\}$$

$$\vec{q} = \vec{t^{\Gamma}_{0}}, \ldots, \vec{t^{\Gamma}_{n}}$$





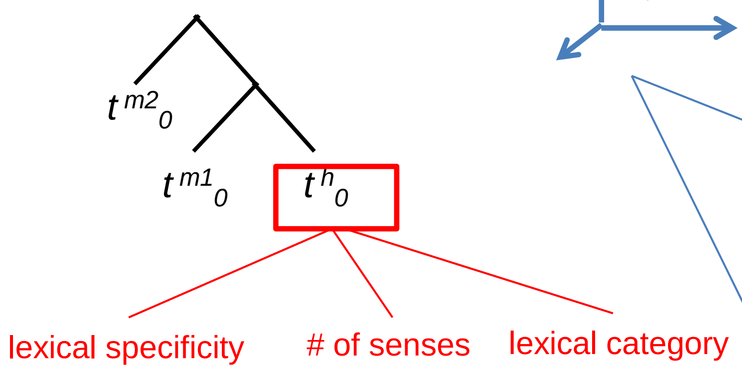
- Vector neighborhood density

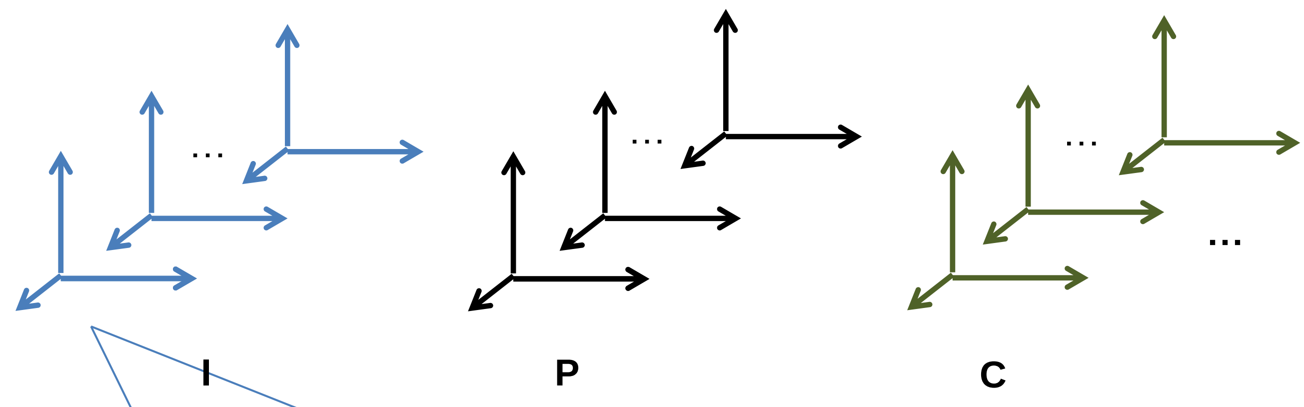
- **Semantic differential**

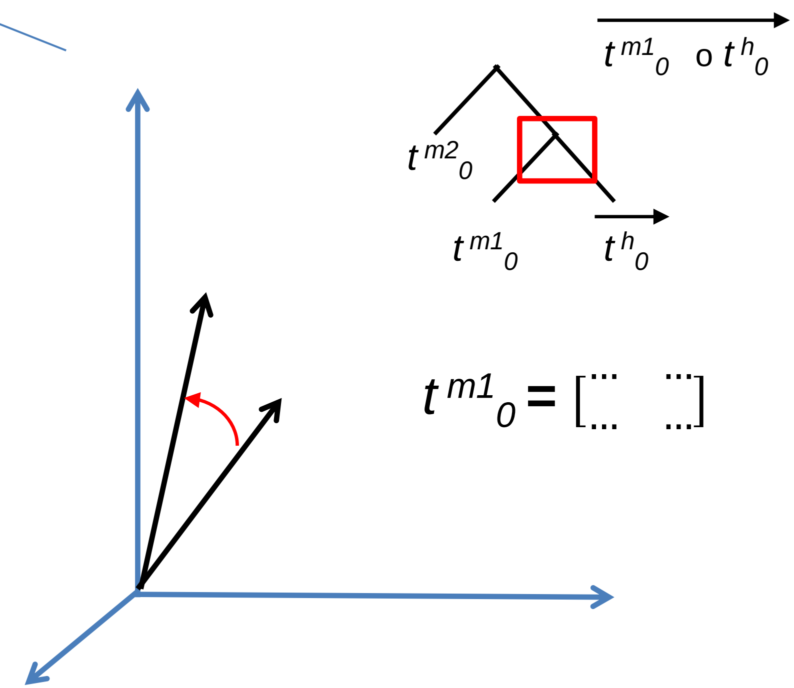
Semantic pivoting

$$\Gamma = \{I, P, C, V\}$$

$$\vec{q} = \overrightarrow{t^{\Gamma}}_0, \dots, \overrightarrow{t^{\Gamma}}_n$$







- Vector neighborhood density

- Semantic differential

- **Distributional compositionality**

Search and Composition Operations

- Instance search
  - Proper nouns
  - String similarity + node cardinality
- Class (unary predicate) search
  - Nouns, adjectives and adverbs
  - String similarity + Distributional semantic relatedness
- Property (binary predicate) search
  - Nouns, adjectives, verbs and adverbs
  - Distributional semantic relatedness

$$sr(\overrightarrow{\mathbf{q}'}_1, \overrightarrow{\mathbf{p}}_0) \geq \eta$$

- Navigation

$$< (\overrightarrow{\mathbf{q}'}_1 - \overrightarrow{\mathbf{p}}_1), (\overrightarrow{\mathbf{q}'}_2 - \overrightarrow{\mathbf{p}}_2), \cdots, (\overrightarrow{\mathbf{q}'}_n - \overrightarrow{\mathbf{p}}_n) >$$

- Extensional expansion
  - Expands the instances associated with a class.
- Operator application
  - Aggregations, conditionals, ordering, position
- Disjunction & Conjunction
- Disambiguation dialog (instance, predicate)

Natural Language Queries over Heterogeneous Linked Data Graphs: A Distributional-Compositional Semantics Approach, IUI 2014

What to expect (@ QALD1)
F1-Score: 0.72
MRR: 0.5

Freitas & Curry, Natural Language Queries over Heterogeneous Linked Data Graphs, IUI (2014).

Software: StarGraph

- Semantic parsing.

https://github.com/Lambda-3/Stargraph

Freitas et al., Natural Language Queries over Heterogeneous Linked Data Graphs: A Distributional-Compositional Semantics Approach, 2014.

Emerging perspectives

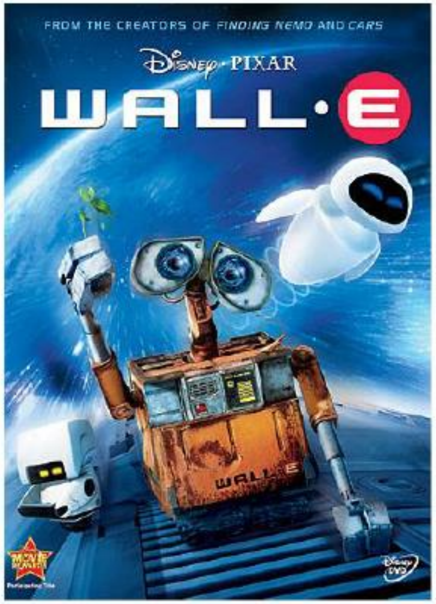
Semantic Parsing:

- Structured queries over KGs as explanations.
- Semantic pivoting heuristics.
- Diversity of distributional/compositional models as key.
- End-to-end vs componentised architectures.

Knowledge Graph Completion

The Problem

WALL-E _has_genre ?



Liu, Representation Learning for Large-Scale Knowledge Graphs, 2015

The Problem

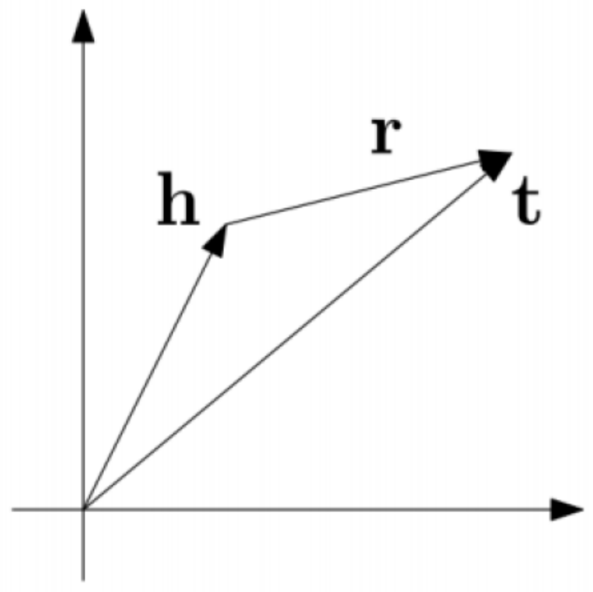
WALL-E _has_genre

- Animation
- Computer animation
- Comedy film
- Adventure film
- Science Fiction
- Fantasy
- Stop motion
- Satire
- Drama
- Connecting

Liu, Representation Learning for Large-Scale Knowledge Graphs, 2015

Formulating the Distributional-Relational Representation

For each triple (head, relation, tail), relation as a translation from head to tail

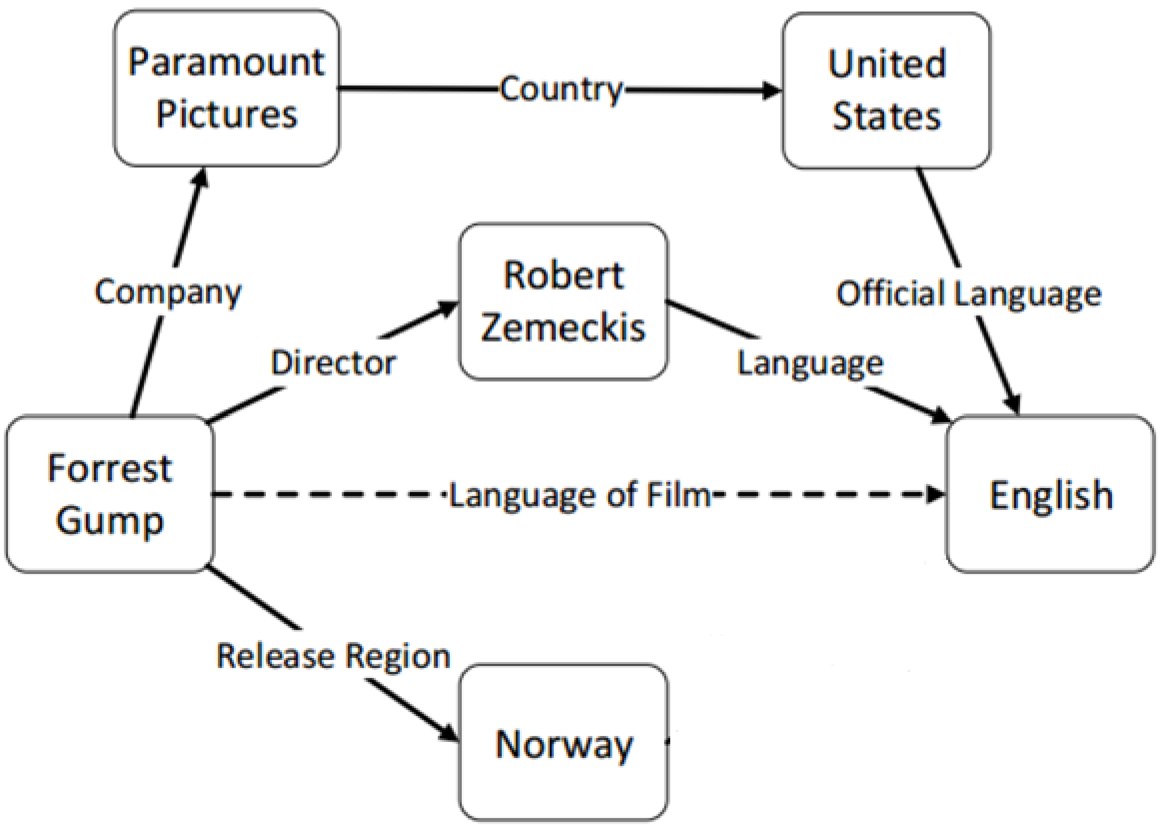


Learning objective: **h + r = t**

Liu, Representation Learning for Large-Scale Knowledge Graphs, 2015

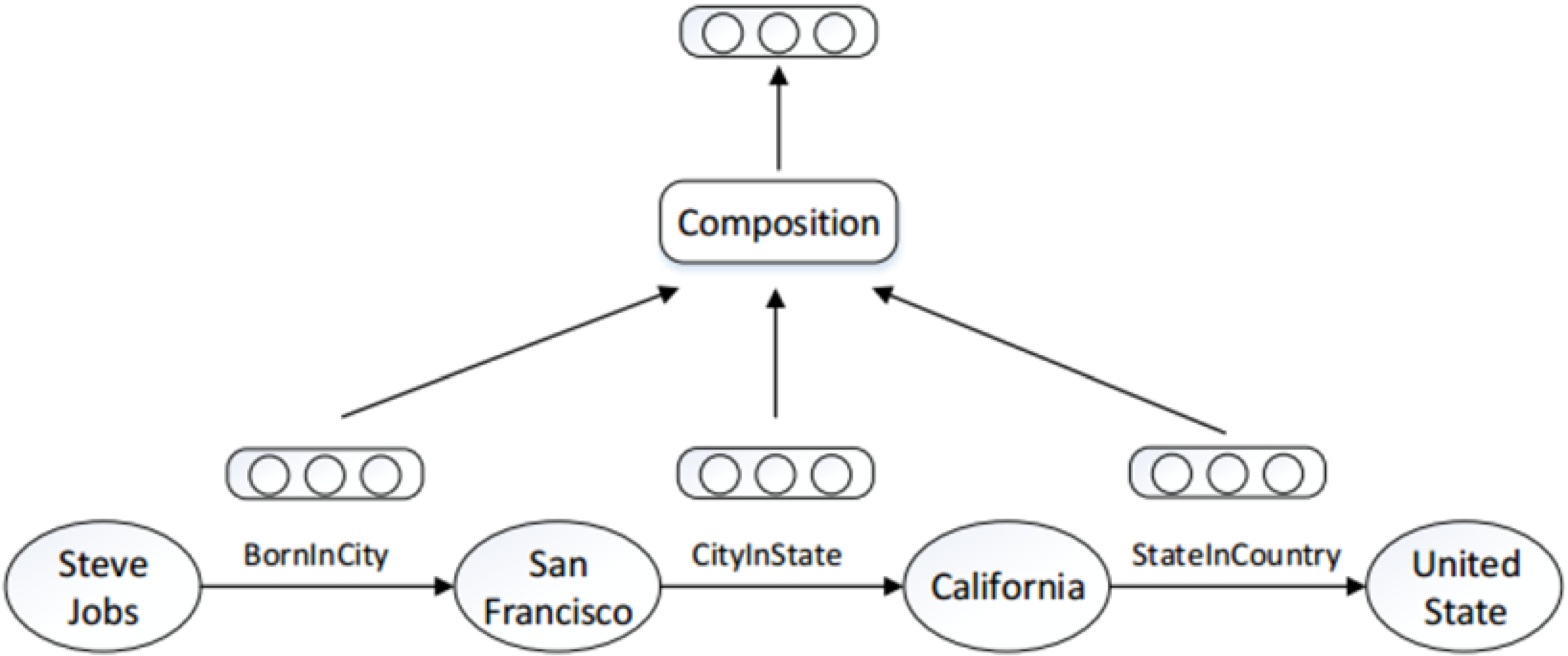
Relation Paths

- Complex Inference patterns for composition.



Liu, Representation Learning for Large-Scale Knowledge Graphs, 2015

Representation of Relation Paths



Gardner, et al. (2013). Improving learning and inference in a large knowledge-base using latent syntactic cues. EMNLP.

Liu, Representation Learning for Large-Scale Knowledge Graphs, 2015

What to expect (PTransE@FB15K)

Relation Prediction

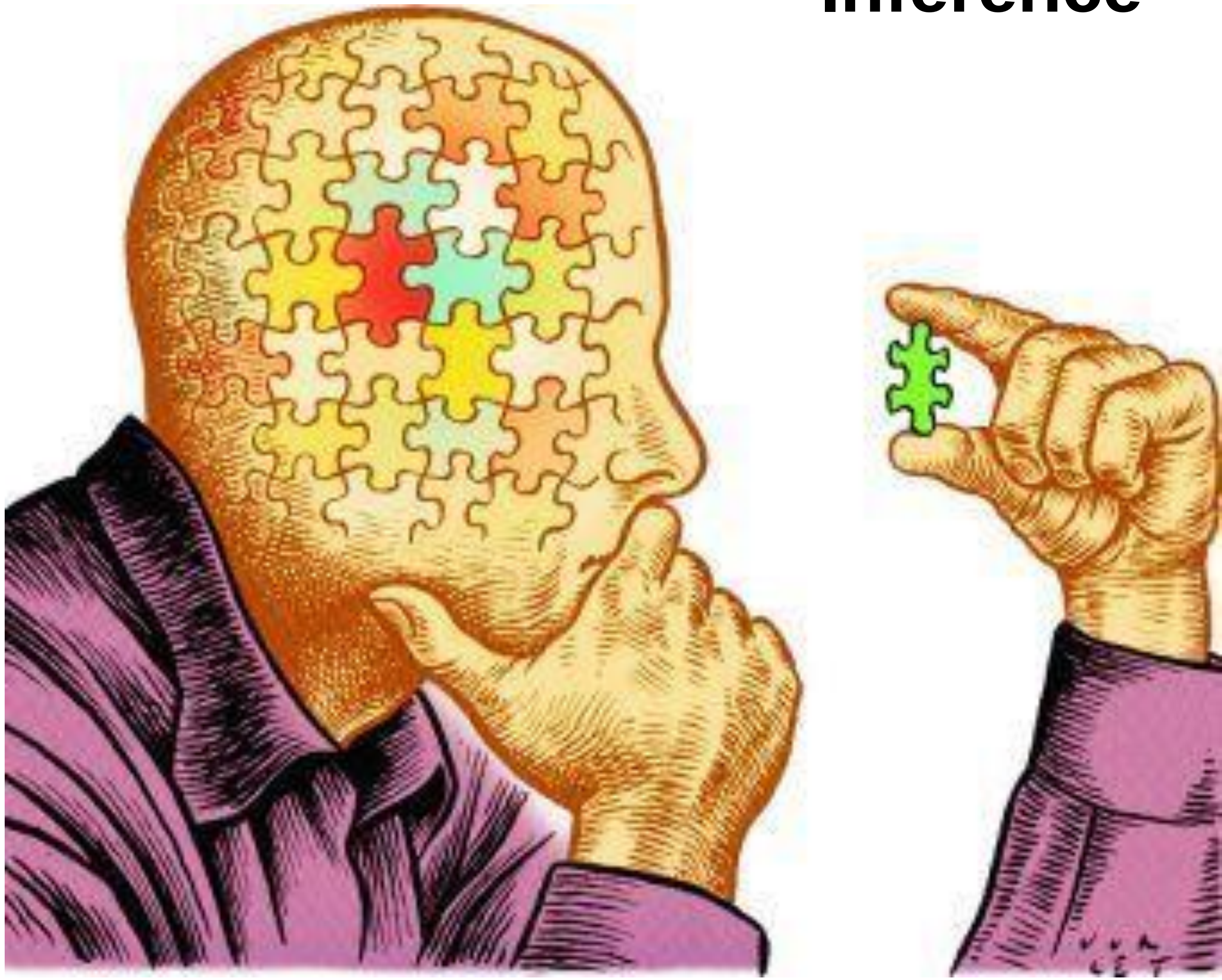
Metric	Mean Rank Raw	Mean Rank Filter	Hits@1 (%) Raw	Hits@1 (%) Filter
TransE	2.8	2.5	65.1	$\boxed{84.3}$
+Rev	2.6	2.3	67.1	86.7
+Rev+Path	2.4	1.9	65.2	89.0
PTransE (ADD, 2-step)	**1.7**	**1.2**	69.5	93.6
-TransE	135.8	135.3	51.4	78.0
-Path	2.0	1.6	**69.7**	89.0
PTransE (MUL, 2-step)	2.5	2.0	66.3	89.0
PTransE (RNN, 2-step)	1.9	1.4	68.3	93.2
PTransE (ADD, 3-step)	1.8	1.4	68.5	$\boxed{\mathbf{94.0}}$

+10%

Lin, et al. (2015). Modeling Relation Paths for Representation Learning of Knowledge Bases. EMNLP.

MANCHESTER
1824
The University of Manchester

Natural Language Inference



MANCHESTER 1824
The University of Manchester

Recognizing and Justifying Text Entailments (TE) using Definition KGs

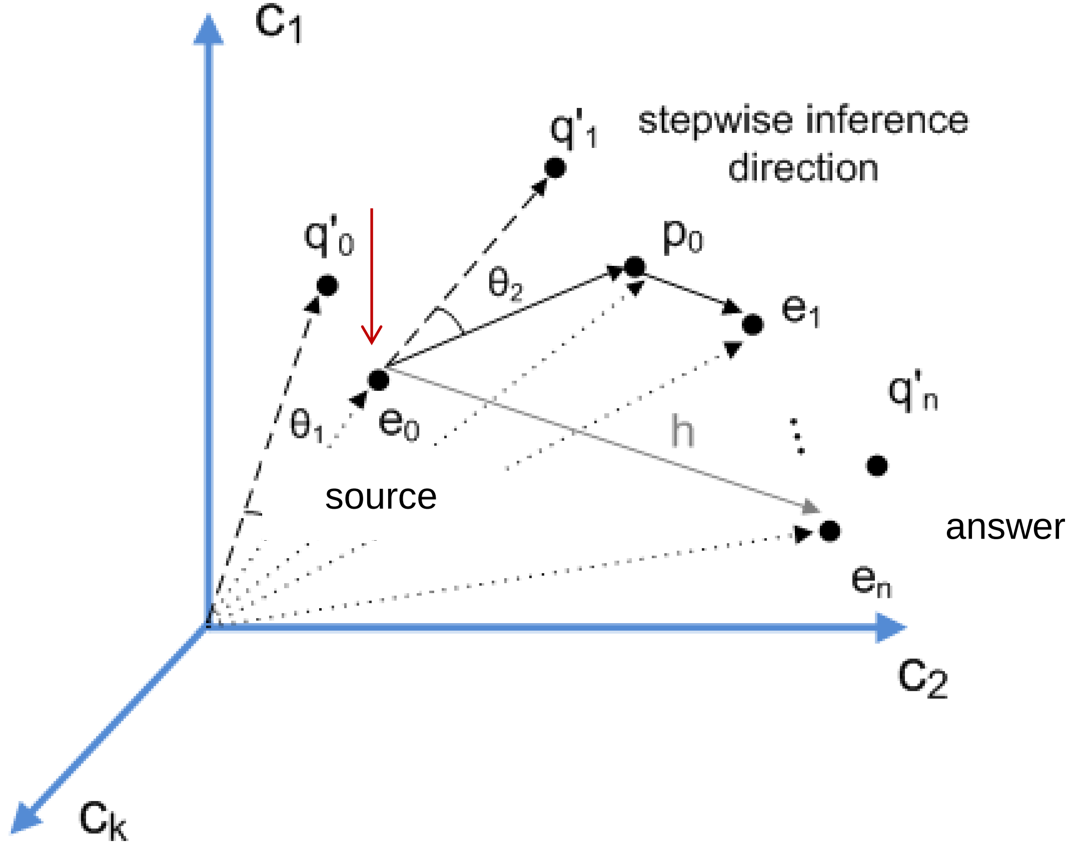
Text Entailment (TE) is a directional relationship between a pair of expressions:

T	the entailing *text*	Many cellphones have built-in digital cameras.
H	the entailed *hypothesis*	Many cellphones can take pictures.

T entails H if a human reading T can infer that H is true

Proposed Approach: besides answering ***if*** the entailment is true, also tell ***why*** it is true

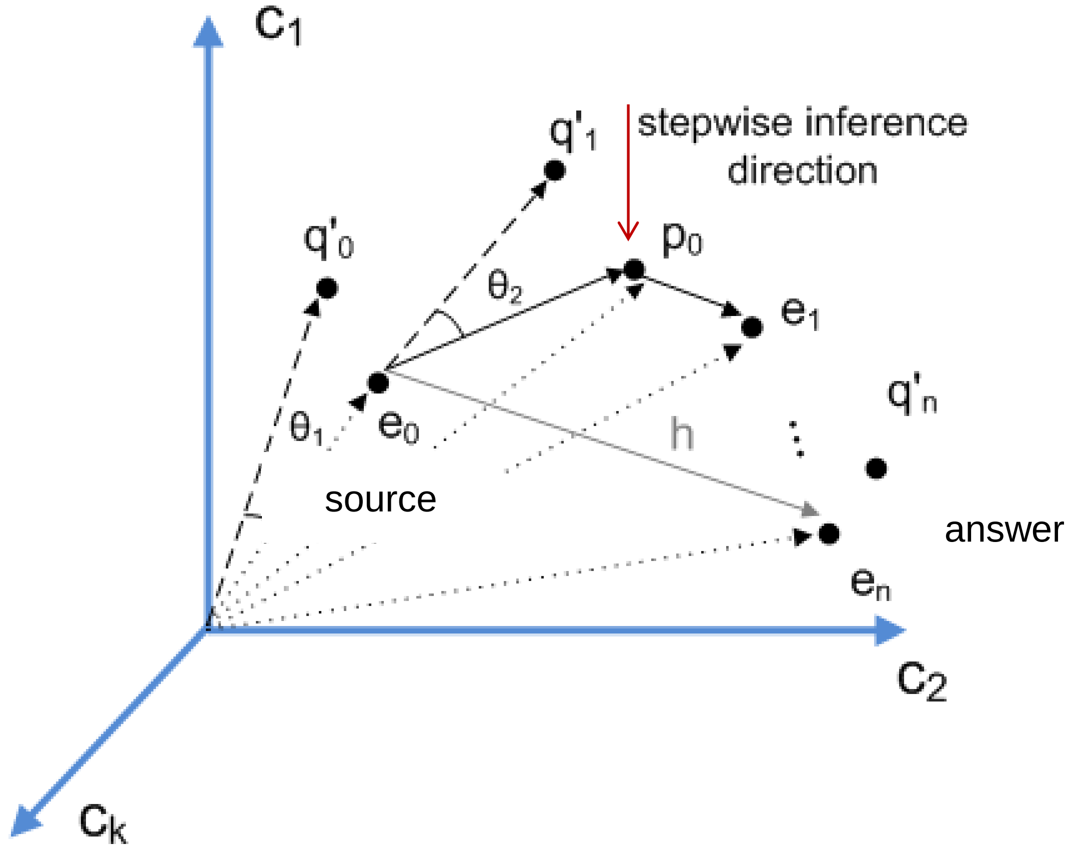
MANCHESTER 1824
The University of Manchester



Distributional heuristics

Distributional semantic relatedness as a Selectivity Heuristics

MANCHESTER 1824
The University of Manchester

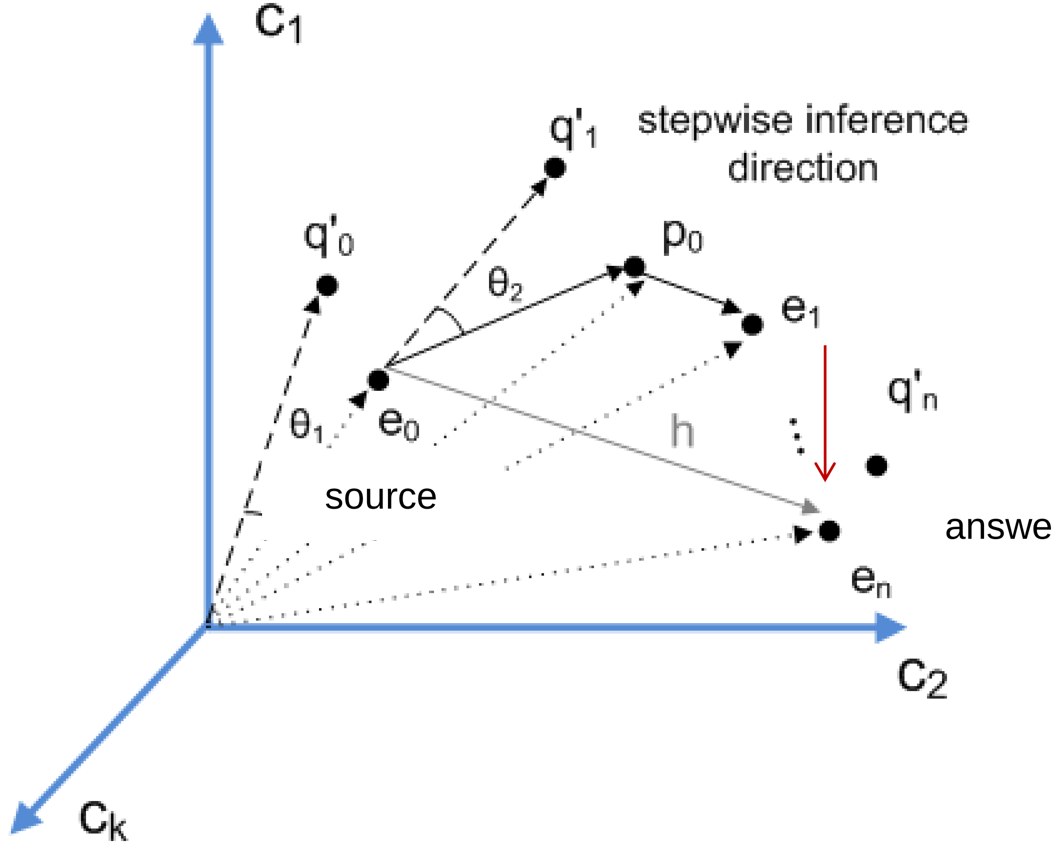


Distributional heuristics

Distributional semantic relatedness as a Selectivity Heuristics

MANCHESTER 1824

The University of Manchester



Distributional heuristics

Distributional semantic relatedness as a Selectivity Heuristics

MANCHESTER 1824
The University of Manchester

Pre-Processing

Pre-processing

1 Text Sentence Simplification

T: Many cellphones have built-in digital cameras, which are very robust.

T_1: Many cellphones have built-in digital cameras.

T_2: Built-in digital cameras are very robust.

H: Many cellphones can take pictures.

edit_dist(T_1, H) < edit_dist(T_2, H)
then the new pair is composed
by ***T_1*** *and* ***H***

2 Core Words and Input Pairs Selection

T: Many cellphones have built-in digital cameras.

discard overlapping words *main verb* *verb complements*

H: Many cellphones can take pictures.

Combine all words, compute the semantic relatedness and select the pairs with the highest scores:

- ✓ [digital cameras, pictures]
- ✓ [digital cameras, take]
- × [have, take]
- × [have, pictures]

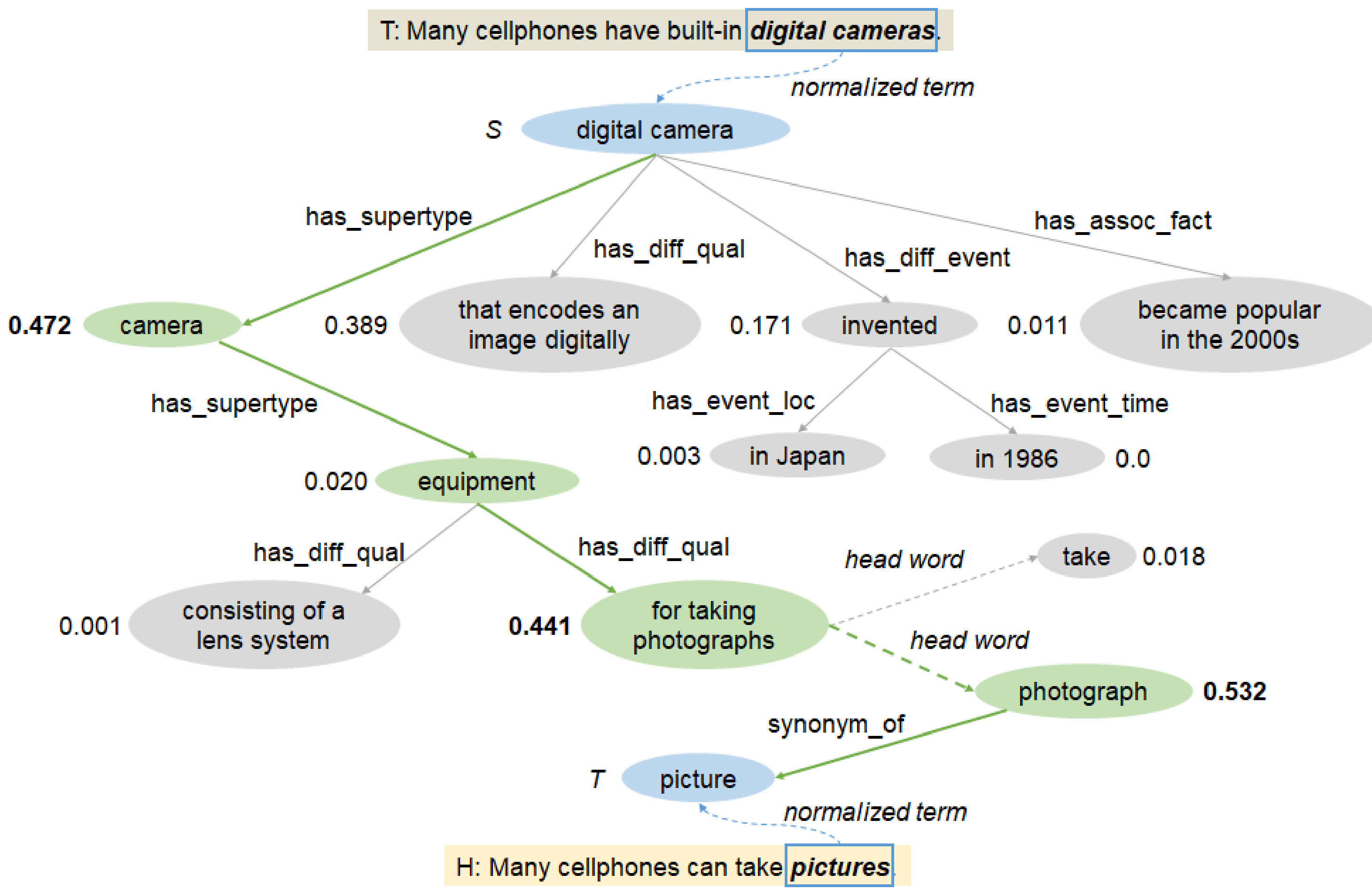
Abductive Inference

The University of Manchester

Reasoning

3 Path Search through DNA on the Definition Graph

Take the pairs found in the previous step as input and perform a depth first search in the graph



Generation

Explanation

4 Explanation Generation

Write the justification, using the text in the nodes in the path found by DNA:

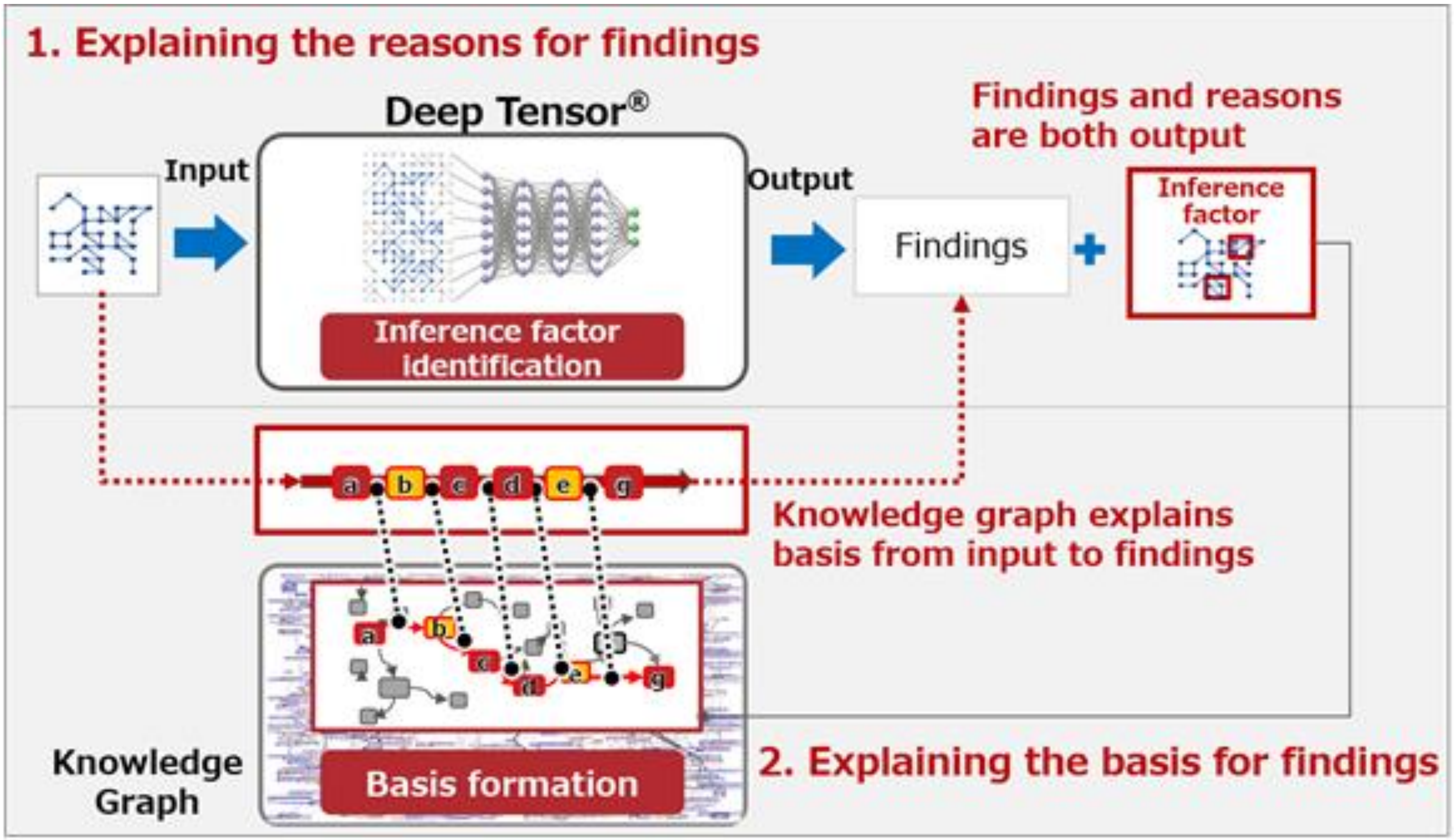
A digital camera is a kind of camera
A camera is an equipment for taking photographs
Photograph is synonym of picture

What to expect (TE@Boeing-Princeton-ISI)
F1-Score: 0.59

What to expect (TE@Guardian Headline Samples)
F1-Score: 0.53

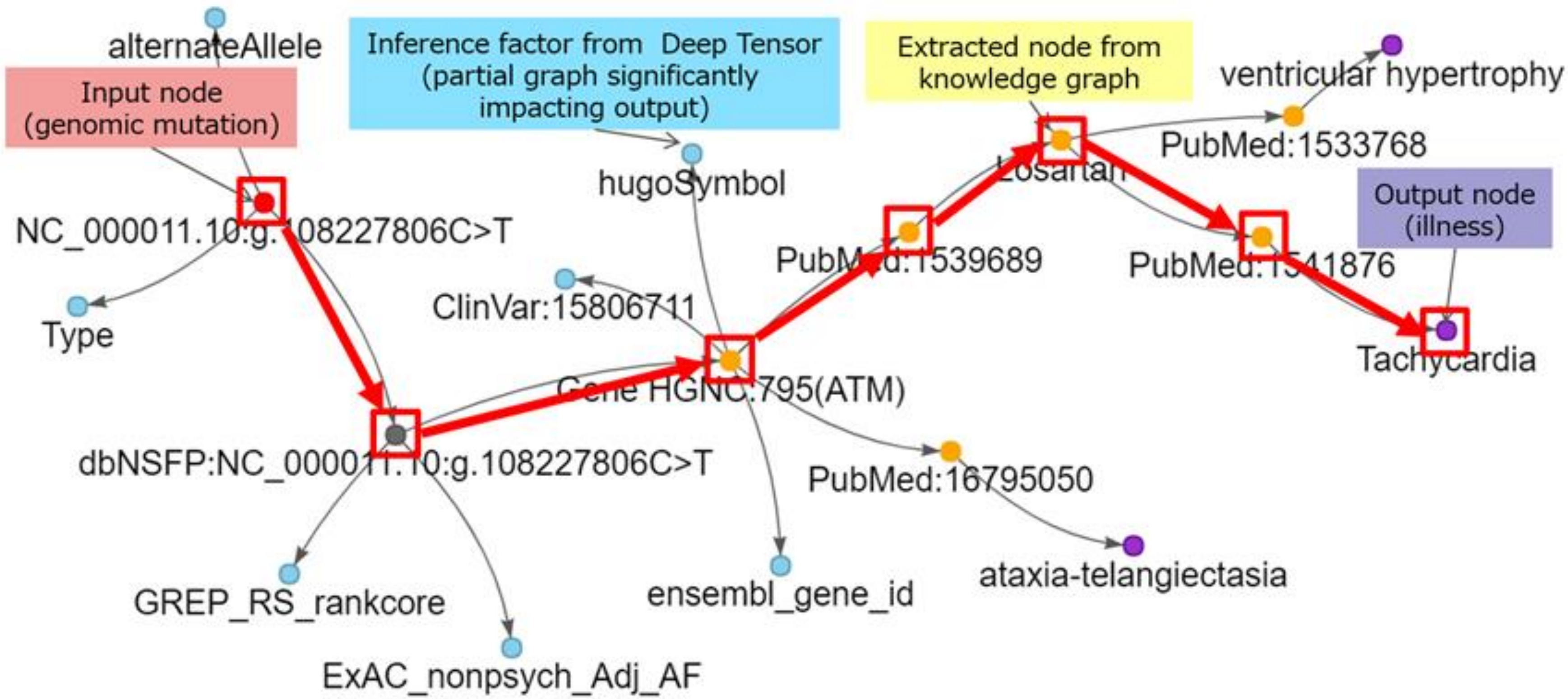
Santos et al., Recognizing and Justifying Text Entailment through Distributional Navigation on Definition Graphs, AAAI, 2018.

Explainable Findings
From Tensor Inferences Back to KGs



Explainable Findings

From Tensor Inferences Back to KGs



MANCHESTER
1824
The University of Manchester

LINKQ
Powered by Thomson Reuters

News Feed | Connection | Portfolio

Settings | All | Pinned

3 number of pinned articles

BRIEF-Avio receives EUR 40 mln financing from European Investment Bank <SPA2.MI>
Companies: Avio SPA **80%**, Avio SPA **80%**, Avio SPA **80%**
Topics: Business Finance, Contracts / Business Deals Events: ContactDetails
Industry: Spacecraft Manufacturing
Publication date: Oct 6, 2017 10:24:02 AM

Airbus SE
BAE Systems PLC
Boeing Co

BRIEF-Airtelis orders three H215 Airbus Helicopters<AIR.PA><AIRG.DE>
Companies: Airbus Helicopters SAS **50%**, Airbus Helicopters SAS **50%**, Airbus Helicopters SAS **50%**
Topics: N/A Events: BusinessRelation
Industry: Aerospace & Defense - NEC, Aircraft Parts Manufacturing - NEC
Publication date: Oct 4, 2017 10:46:00 AM

Raytheon Co
BAE Systems PLC
Boeing Co

BRIEF-British Airline Pilots' Association - pilots union calls for investigation into collapse of Monarch Airlines<MONA.UL>
Companies: Monarch Airlines Ltd **80%**, Monarch Airlines Ltd **80%**, Monarch Airlines Ltd **80%**
Topics: Other Events: N/A
Industry: Airlines - NEC
Publication date: Oct 9, 2017 1:55:13 PM

Raytheon Co
BAE Systems PLC
Boeing Co

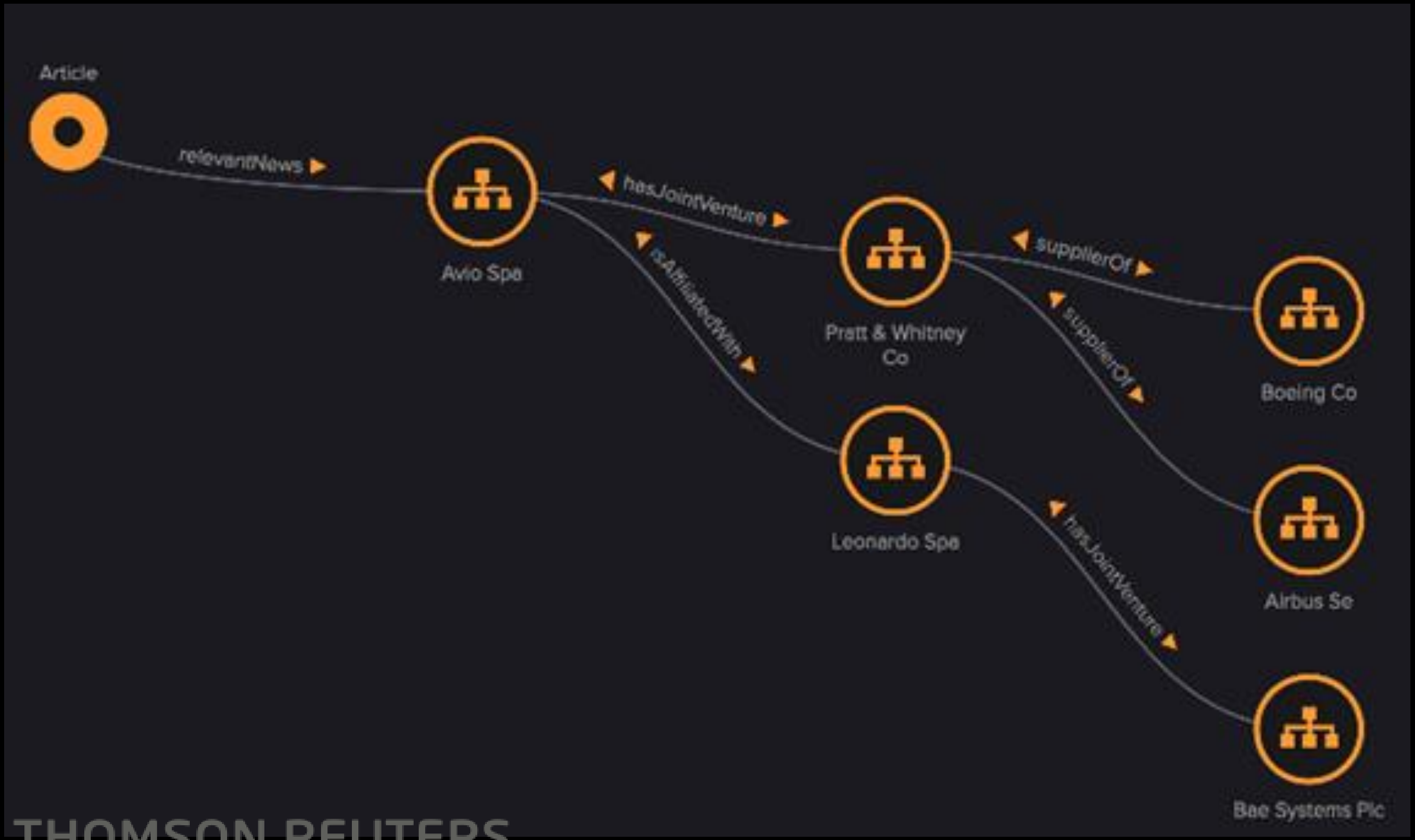
THOMSON REUTERS

MANCHESTER 1824

The University of Manchester

BRIEF-Avio receives EUR 40 mln financing from European Investment Bank <SPA2.MI>

Oct 6 (Reuters) - AVIO SPA <SPA2.MI> * SAYS SIGNED WITH EUROPEAN INVESTMENT BANK CONTRACT FOR EUR 40 MILLION FINANCING Source text for Eikon: [ID:nBIA5D9Ntk] Further company coverage: [SPA2.MI] (Gdynia Newsroom) ((gdynia.newsroom@thomsonreuters.com; +48 58 772 0920 ;))



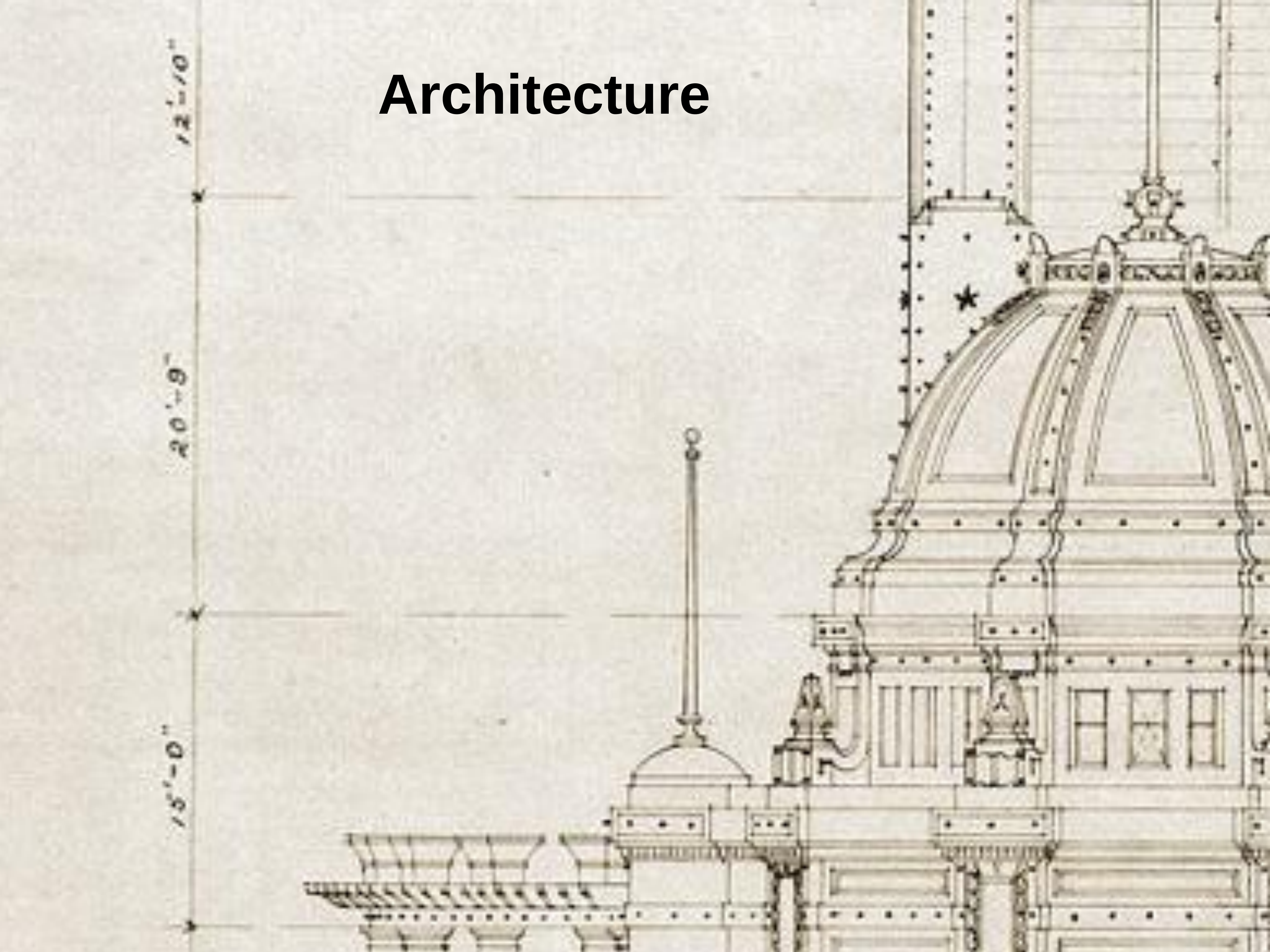
THOMSON REUTERS

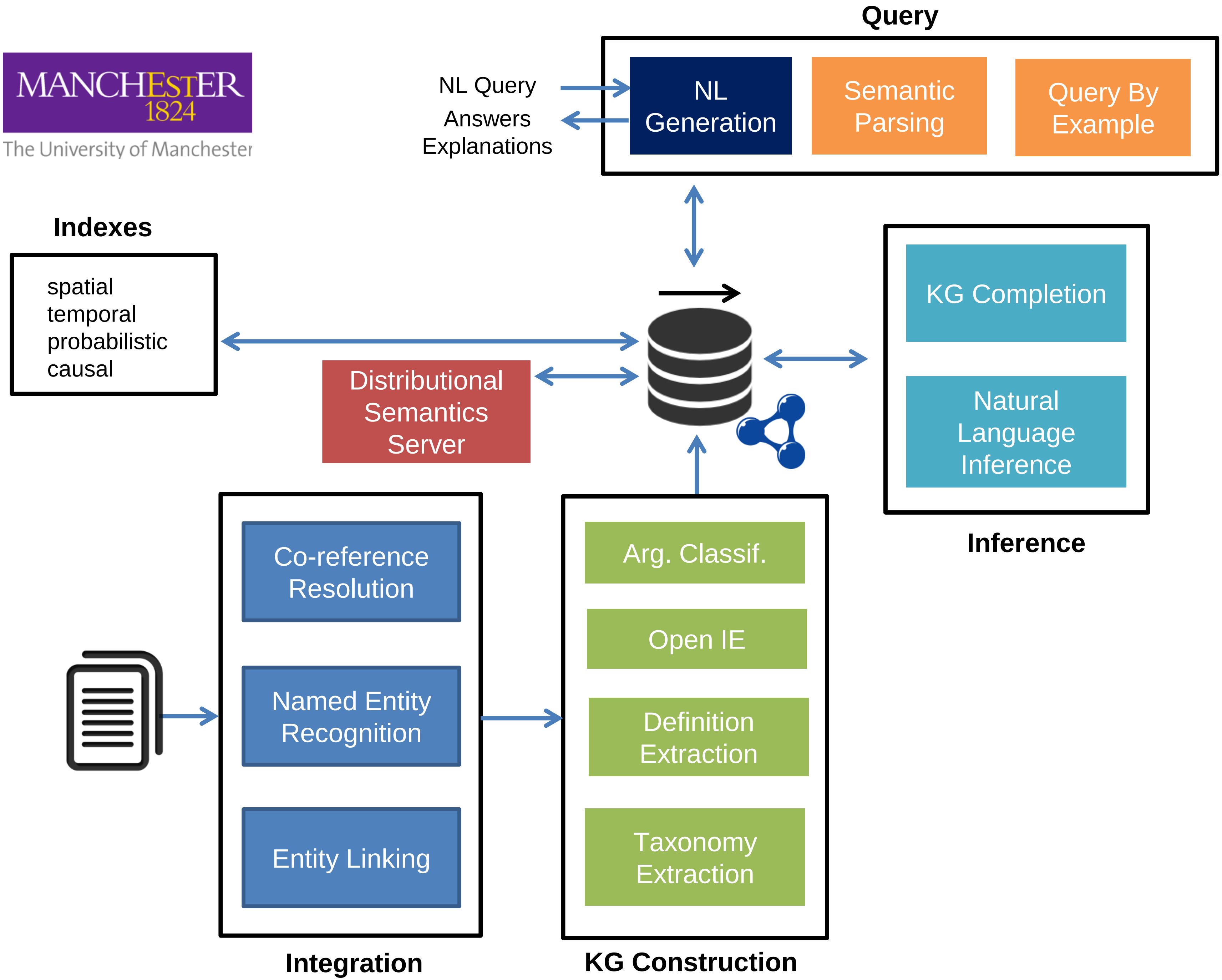
Emerging perspectives

- Distributional-relational models in KB completion explored a large range of representation paradigms.
  - Opportunity for exporting these representation models to other tasks.

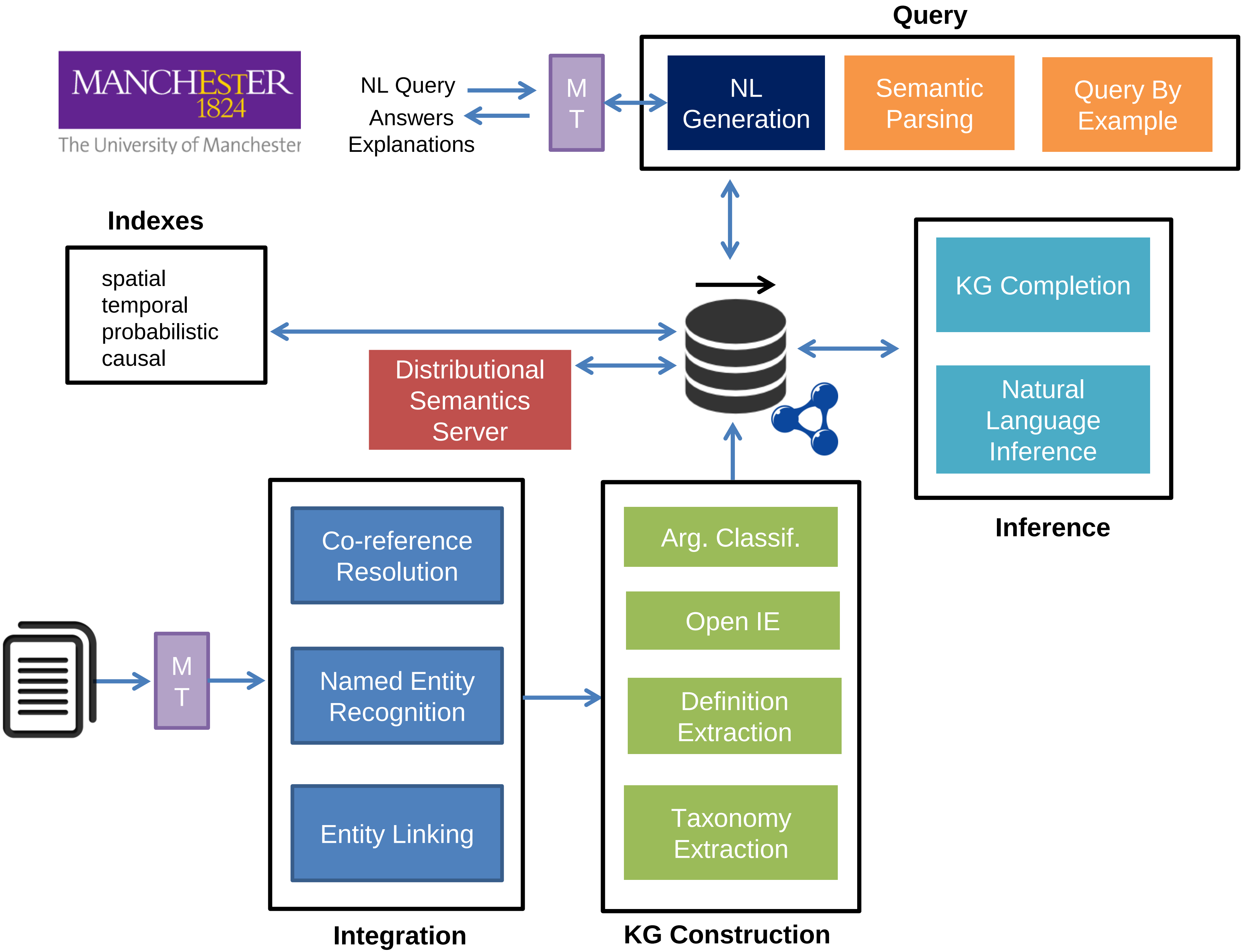
- Definition-based models can provide a corpus-viable, low-data and explainable alternative to embedding-based models.

Architecture





Query
NL Query
Answers
Explanations
NL Generation
Semantic Parsing
Query By Example
Indexes
spatial
temporal
probabilistic
causal
Distributional Semantics Server
KG Completion
Natural Language Inference
Inference
Co-reference Resolution
Named Entity Recognition
Entity Linking
Integration
Arg. Classif.
Open IE
Definition Extraction
Taxonomy Extraction
KG Construction



MANCHESTER 1824
The University of Manchester
NL Query
Answers
Explanations
M
T
Query
NL Generation
Semantic Parsing
Query By Example
Indexes
spatial
temporal
probabilistic
causal
Distributional Semantics Server
KG Completion
Natural Language Inference
Inference
Co-reference Resolution
Named Entity Recognition
Entity Linking
Integration
Arg. Classif.
Open IE
Definition Extraction
Taxonomy Extraction
KG Construction
M
T

**Take
Home
Message**

Take Home Message

- The evolution of methods, tools and the availability of data in NLP creates the demand for knowledge representation models to support complex AI systems.

- A relaxed version of RDF (RDF-NL) can provide this answer.
  - Establishes a dialogue with a standard (with existing data).
  - Inherits optimization aspects from Databases.

- Word-vectors (DSMs) + compositional models + RDF-NL.

- Moving beyond facts and taxonomies: rhetorical structures, arguments, polarity, stories.

Take Home Message

- Syntactic and lexical features can go a long way for structuring text.
  - Context-preserving open information extraction.
- Integration (entity reconciliation) as semantic-best effort.
  - Embrace schema on read.
- KGs can support explainable AI:
  - Meeting point between extraction, reasoning and querying.
  - Definition-based models.
- Inherit infrastructures from DB and IR.

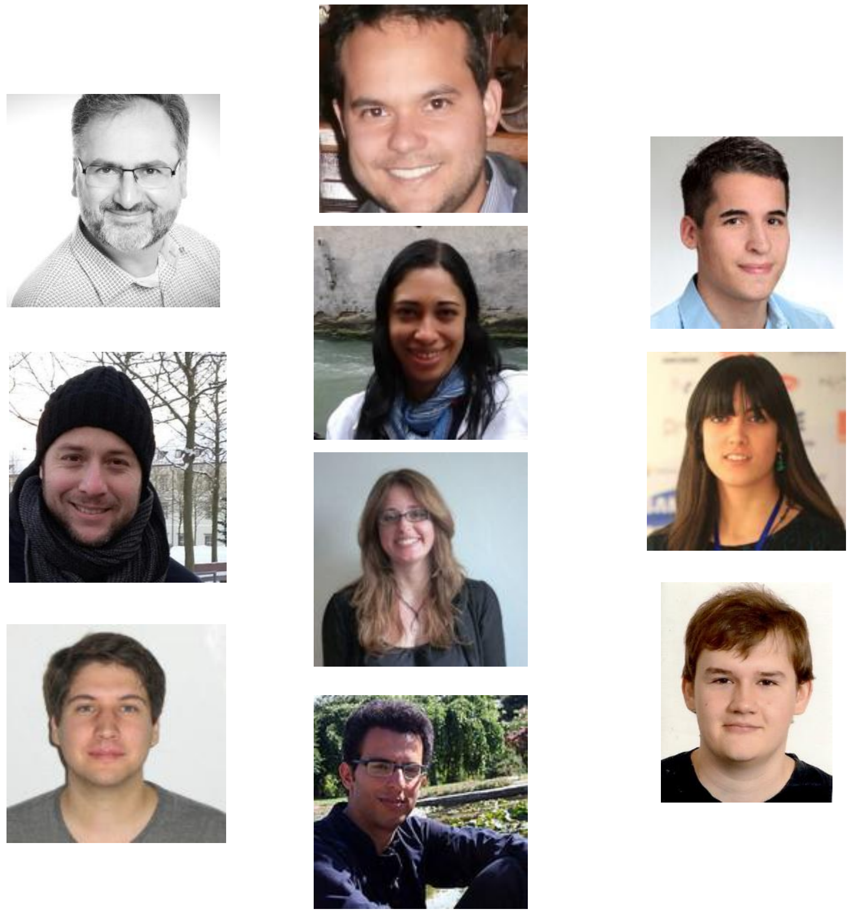
Take Home Message

Opportunities:

- ML orchestrated pipelines with:
  - Richer discourse-representation models.
  - Explicit semantic representations (centered on KGs).
  - Different compositional/distributional models (beyond W2V & Glove)
- KGs and impact on explainability.
- Quantifying domain and language transportability.

MANCHESTER
1824
The University of Manchester

Acknowledgements



The University of Manchester

References

Vivian S. Silva, André Freitas, Siegfried Handschuh, *Recognizing and Justifying Text Entailment through Distributional Navigation on Definition Graphs*, Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, USA, 2018 (pdf).

Matthias Cetto, Christina Niklaus, André Freitas and Siegfried Handschuh, *Creating a Hierarchy of Semantically-Linked Propositions in Open Information Extraction*, In Proceedings of the 27th International Conference on Computational Linguistics (COLING), New-Mexico, USA, 2018.

Christina Niklaus, Matthias Cetto, André Freitas and Siegfried Handschuh, *A Survey on Open Information Extraction*, In Proceedings of the 27th International Conference on Computational Linguistics (COLING), New-Mexico, USA, 2018.

Siamak Barzegar, Brian Davis, Siegfried Handschuh, André Freitas, *Multilingual Semantic Relatedness using Lightweight Machine Translation*, 12th IEEE International Conference on Semantic Computing (ICSC), USA, 2018 (pdf).

Siamak Barzegar, Brian Davis,, Manel Zarrouk, Siegfried Handschuh, André Freitas, *SemR-11: A Multi-Lingual Gold-Standard for Semantic Similarity and Relatedness for Eleven Languages*, 11th Language Resources and Evaluation Conference (LREC), Japan, 2018 (pdf).

Thomas Gaillat, Manel Zarrouk, André Freitas, Brian Davis, *The SSIX Corpora: Three Gold Standard Corpora for Sentiment Analysis in English, Spanish and German Financial Microblogs,* 11th Language Resources and Evaluation Conference (LREC), Japan, 2018 (pdf).

Juliano Efson Sales, Leonardo Souza, Siamak Barzegar, Brian Davis, André Freitas and Siegfried Handschuh, *Indra: A Word Embedding and Semantic Relatedness Server*, 11th Language Resources and Evaluation Conference (LREC), Japan, 2018 (pdf).

Andre Freitas, Schema-agnostic queries over large-schema databases: a distributional semantics approach (pdf).

References

Vivian S. Silva, André Freitas, Siegfried Handschuh, *Building a Knowledge Graph from Natural Language Definitions for Interpretable Text Entailment Recognition*, 11th Language Resources and Evaluation Conference (LREC), Japan, 2018 (pdf).

Juliano Efson Sales, André Freitas, Brian Davis, Siegfried Handschuh, *A Compositional-Distributional Semantic Model for Searching Complex Entity Categories*, 5th Joint Conference on Lexical and Computational Semantics (*SEM), Berlin, 2016. (Full Conference Paper) (pdf).

André Freitas, Siamak Barzegar, Juliano E. Sales, Siegfried Handschuh and Brian Davis, *Semantic Relatedness for All (Languages): A Comparative Analysis of Multilingual Semantic Relatedness using Machine Translation*, 20th International Conference on Knowledge Engineering and Knowledge Management (EKAW), 2016. (Full Conference Paper) (pdf).

Vivian S. Silva, André Freitas and Siegfried Handschuh, *Supersense Word Tagging with Foundational Ontology Classes, 20th International Conference on Knowledge Engineering and Knowledge Management* (EKAW), 2016. (Full Conference Paper) (pdf).

Christina Niklaus, Bernhard Bermeitinger, Siegfried Handschuh, André Freitas, *A Sentence Simplification System for Improving Relation Extraction*, 26th International Conference on Computational Linguistics, (COLING), Osaka, 2016. (Demonstration in Proceedings) (pdf).

Vivian S. Silva, Siegfried Handschuh and André Freitas, *Categorization of Semantic Roles for Dictionary Definitions*, Cognitive Aspects of the Lexicon (CogALex-V), Workshop at the 26th International Conference on Computational Linguistics, (COLING), Osaka, 2016. (Full Workshop Paper) (pdf).

André Freitas, Juliano Efson Sales, Siegfried Handschuh, Edward Curry, *How hard is the Query? Measuring the Semantic Complexity of Schema-Agnostic Queries*, In Proceedings of the 11th International Conference on Computational Semantics (IWCS), London, 2015. (Full Conference Paper) (pdf).

André Freitas, Edward Curry, *Natural Language Queries over Heterogeneous Linked Data Graphs: A Distributional-Compositional Semantics Approach*, In Proceedings of the 19th International Conference on Intelligent User Interfaces (IUI), Haifa, 2014. (Full Conference Paper) (pdf).

André Freitas, João Carlos Pereira Da Silva, Edward Curry, Paul Buitelaar, *A Distributional Semantics Approach for Selective Reasoning on Commonsense Graph Knowledge Bases*, In Proceedings of the 19th International Conference on Applications of Natural Language to Information Systems (NLDB), Montpellier, 2014. (Full Conference Paper) (pdf).