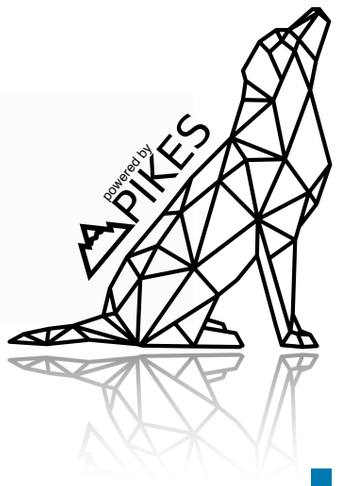


KE4IR

Knowledge Extraction for Information Retrieval

Marco Rospocher

FONDAZIONE BRUNO KESSLER

rospocher@fbk.eu
dkm.fbk.eu/rospocher
@marcorospocher

joint work with:

Francesco Corcoglioniti, Mauro Dragoni, Alessio Palmero Aprosio

13th ESWC 2016 | 29 May - 2 June 2016 | Anissaras, Crete, Greece

Main Message

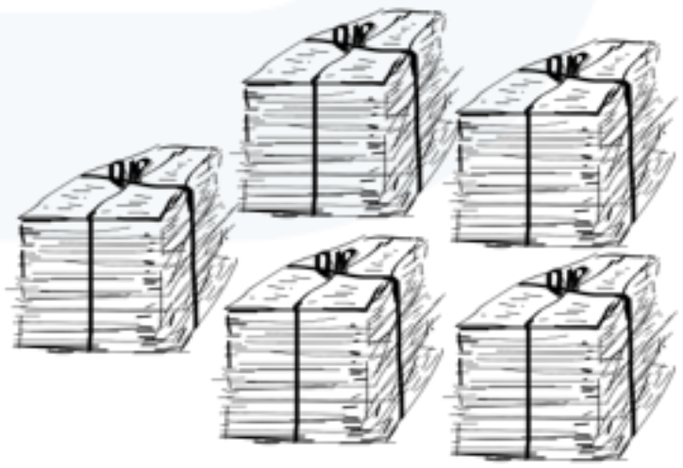
- Exploiting the knowledge extracted from
  - queries
  - documents

  improves Document Retrieval performances!

Outline

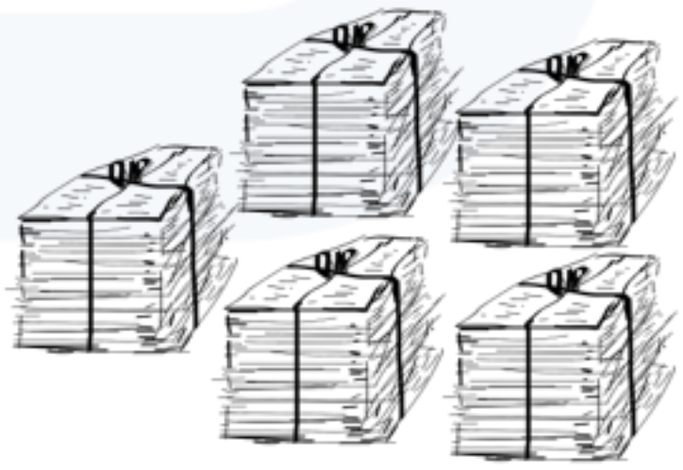
- Document Retrieval and Motivation
- Our approach: KE4IR
- Evaluation: Results and Findings

Document Retrieval



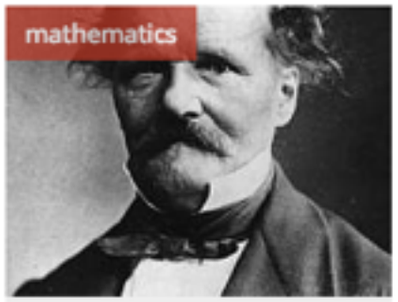
Document Retrieval

astronomers influenced by Gauss



Document Retrieval

astronomers influenced by Gauss



Ernst Kummer and his Achievements in Mathematics

29. January 2015 0 Harald Sack

On January 29, 1810, German mathematician Ernst Eduard Kummer was born. One of his major contributions is the introduction of ideal numbers, which are defined as a special subgroup of a ring, extended the fundamental theorem of arithmetic to complex number fields.He also discovered the fourth order surface based on...

Heinrich Olbers and the Olbers' Paradox

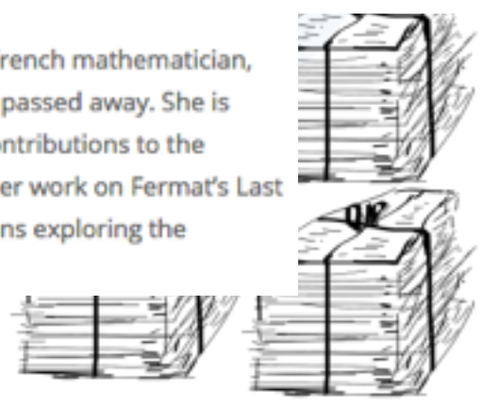
11. October 2014 0 Harald Sack

Heinrich Olbers (1758-1840) On October 11, 1758, German physician and astronomer Heinrich Wilhelm Matthias Olbers was born. Besides his discovery of coments and minor planets, Olbers is best known for his new method to calculate the velocity of falling stars. Maybe you have also heard of the famous Olbers' paradox, which asks...

Sophie Germain and the Chladni Experin

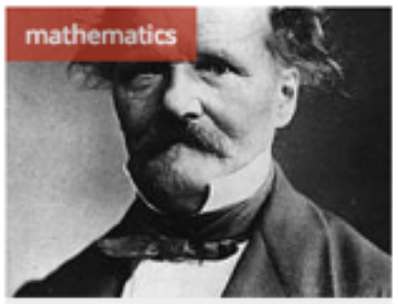
27. June 2014 0 Tabea Tietz

Sophie Germain (1776 – 1831) On June 27, 1831, French mathematician, physicist, and philosopher Marie-Sophie Germain passed away. She is best known for her work in number theory and contributions to the applied mathematics of acoustics and elasticity. Her work on Fermat's Last Theorem provided a foundation for mathematicians exploring the subject...



Document Retrieval

astronomers influenced by Gauss



Ernst Kummer and his Achievements in Mathematics

29. January 2015 [illegible] Tack

On January 29, 1810, [illegible]atician Ernst Eduard Kummer was born. One of his maj[illegible] the introduction of ideal numbers, which are defined as a[illegible]up of a ring, extended the fundamental theorem of arithmetic to complex number fields.He also discovered the fourth order surface based on...

2

Heinrich Olbers and the Olbers' Paradox

11. Oct[illegible] 0 Harald Sack

Heinri[illegible]840) On October 11, 1758, German physician and astron[illegible]helm Matthias Olbers was born. Besides his discove[illegible]nd minor planets, Olbers is best known for his new method to calculate the velocity of falling stars. Maybe you have also heard of the famous Olbers' paradox, which asks...

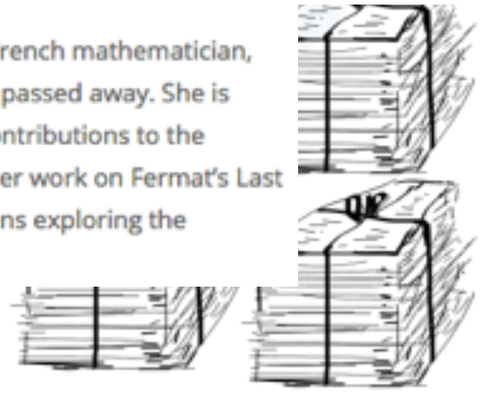
1

Sophie Germain and the Chladni Experin

27. June 2014 0 [illegible]bea Tietz

Sophie Germain[illegible] June 27, 1831, French mathematician, physicist, and p[illegible]Sophie Germain passed away. She is best known for [illegible]er theory and contributions to the applied mathemat[illegible]ics and elasticity. Her work on Fermat's Last Theorem provided a foundation for mathematicians exploring the subject...

3



Motivation

Overcome limitations of traditional IR

- Traditional IR systems match the terms or possible term-based expansions (e.g., synonyms, related terms)

- Issues:
  - relevant documents may not contain all the query terms
  - a document having all terms is not necessarily highly relevant

KE4IR

In a nutshell

- Complement textual terms with semantic terms extracted from queries and documents

KE4IR

In a nutshell

- Complement textual terms with semantic terms extracted from queries and documents

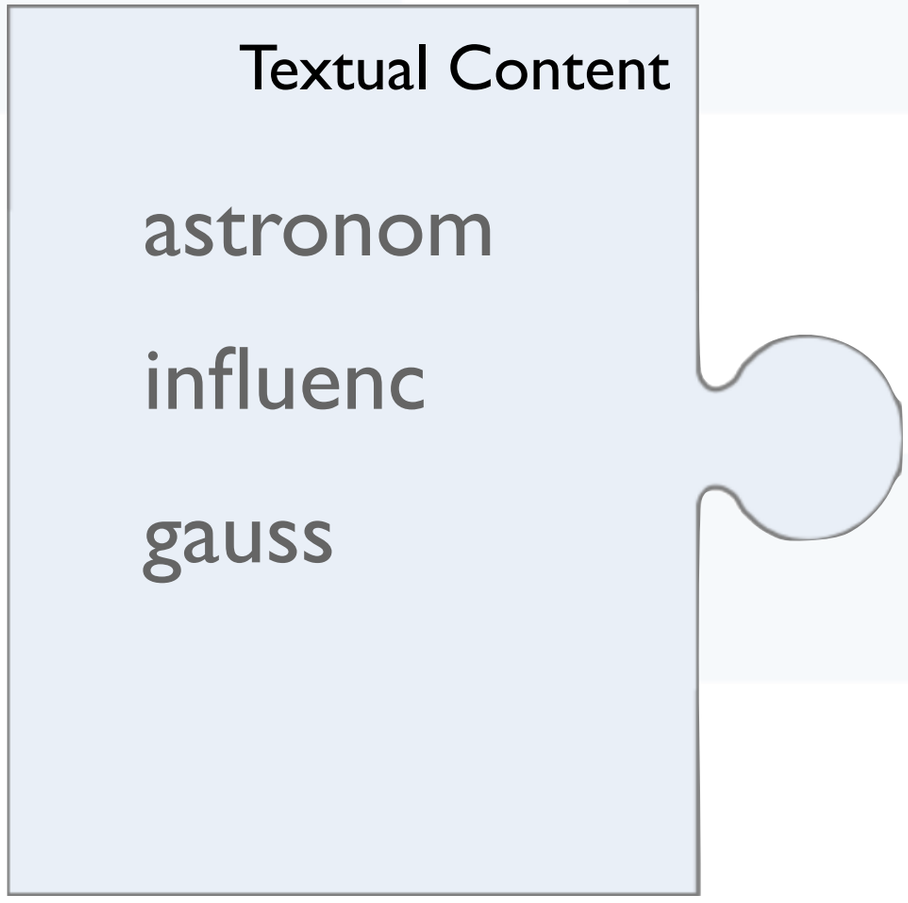
astronomers influenced by Gauss

KE4IR

In a nutshell

- Complement textual terms with semantic terms extracted from queries and documents

astronomers influenced by Gauss

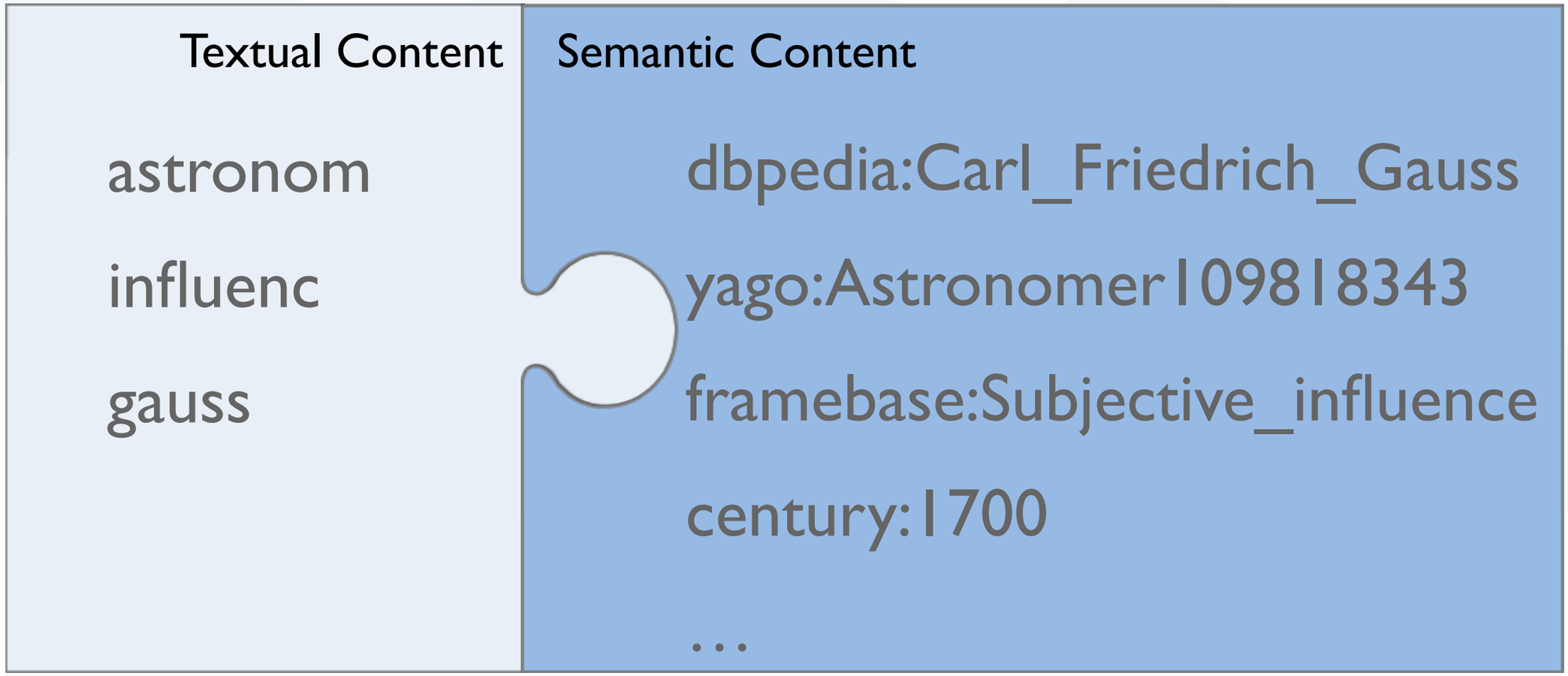


KE4IR

In a nutshell

- Complement textual terms with semantic terms extracted from queries and documents

astronomers influenced by Gauss



KE4IR

Mentions & Semantic Terms

- Mentions are snippets of text denoting entities, events and relations

astronomers influenced by Gauss

- One mention ⇨ A set of semantic terms

- Relevance of a semantic term: number of mentions a term derives from

KE4IR

Mentions & Semantic Terms

- Mentions are snippets of text denoting entities, events and relations

astronomers influenced by Gauss

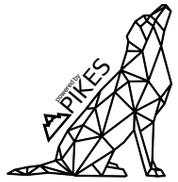
- One mention ⇨ A set of semantic terms

- Relevance of a semantic term: number of mentions a term derives from

KE4IR

Semantic Layers - 1. URI (aka “entities”)

astronomers influenced by Gauss

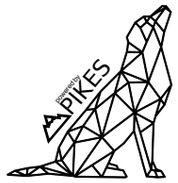


KE4IR

Semantic Layers - 1. URI (aka “entities”)

astronomers influenced by Gauss

dbpedia:Carl_Friedrich_Gauss



KE4IR

Semantic Layers - 2. TYPES

astronomers influenced by Gauss

KE4IR

Semantic Layers - 2. TYPES

astronomers influenced by Gauss

yago:Astronomer109818343

KE4IR

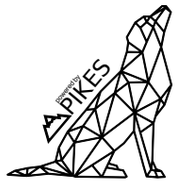
Semantic Layers - 2. TYPES

astronomers influenced by Gauss

yago:Astronomer109818343

dbpedia:Carl_Friedrich_Gauss

yago:GermanMathematicians



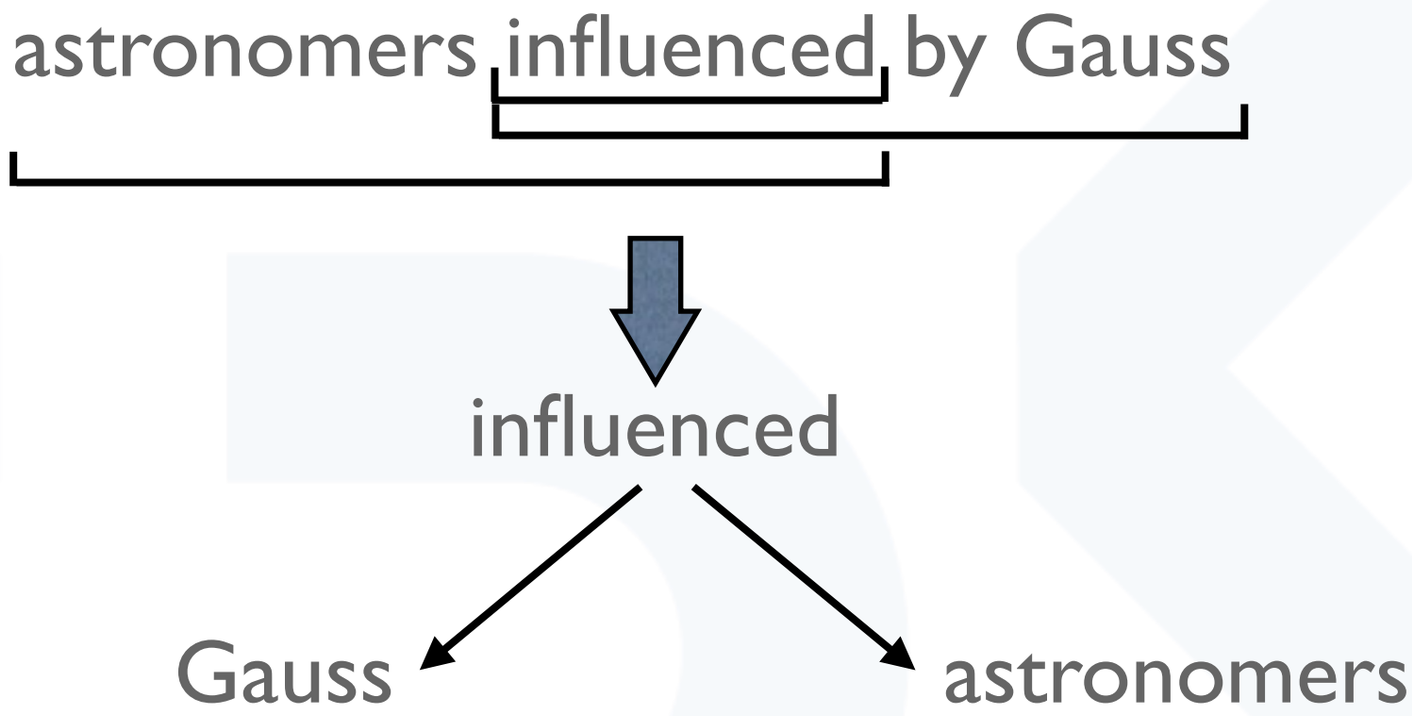
KE4IR

Semantic Layers - 3. FRAME

astronomers influenced by Gauss

KE4IR

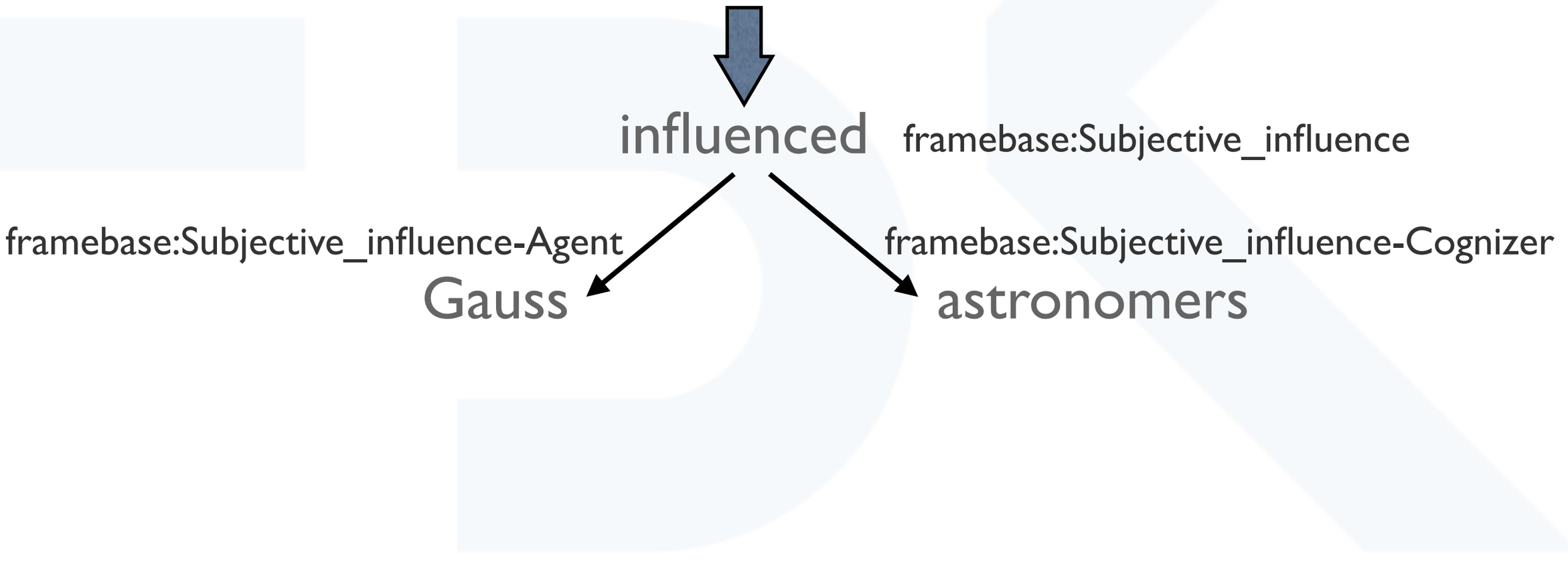
Semantic Layers - 3. FRAME



KE4IR

Semantic Layers - 3. FRAME

astronomers influenced by Gauss



KE4IR

Semantic Layers - 3. FRAME

astronomers influenced by Gauss

influenced framebase:Subjective_influence

framebase:Subjective_influence-Agent

Gauss

framebase:Subjective_influence-Cognizer

astronomers

<framebase:Subjective_influence , dbpedia:Carl_Friedrich_Gauss>

KE4IR

Semantic Layers - 4. TIME

astronomers influenced by Gauss

KE4IR

Semantic Layers - 4. TIME

astronomers influenced by Gauss

dbpedia:Carl_Friedrich_Gauss

dbo:dateOfBirth “1777”

KE4IR

Semantic Layers - 4. TIME

XVIII century astronomers influenced by Gauss

dbpedia:Carl_Friedrich_Gauss

dbo:dateOfBirth “1777”

KE4IR

Semantic Layers - 4. TIME

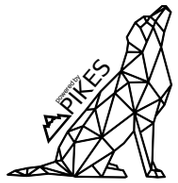
XVIII century astronomers influenced by Gauss

century:18

dbpedia:Carl_Friedrich_Gauss

dbo:dateOfBirth “1777”

year:1777 decade:177 century:17

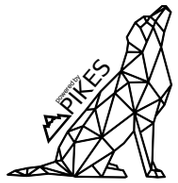


KE4IR

Summing Up

astronomers influenced by Gauss

Layer	Term	Mentions
TEXTUAL	astronom	astronomers
TEXTUAL	influenc	influenced
TEXTUAL	gauss	Gauss



KE4IR

Summing Up

astronomers influenced by Gauss

Layer	Term	Mentions
TEXTUAL	astronom	astronomers
TEXTUAL	influenc	influenced
TEXTUAL	gauss	Gauss
URI	dbpedia:Carl Friedrich Gauss	Gauss
TYPE	yago:GermanMathematicians	Gauss
TYPE	yago:NumberTheorists	Gauss
TYPE	yago:FellowsOfTheRoyalSociety	Gauss
TYPE	*...other 18 terms ...*	Gauss
TYPE	yago:Astronomer109818343	astronomers, Gauss
TYPE	yago:Physicist110428004	astronomers, Gauss
TYPE	yago:Person100007846	astronomers, Gauss
TYPE	*...other 9 terms ...*	astronomers, Gauss
FRAME	⟨Subjective influence-influence.v, Carl . . . Gauss⟩	influenced
FRAME	⟨Subjective influence, Carl Friedrich Gauss⟩	influenced
FRAME	⟨Frame, Carl Friedrich Gauss⟩	influenced
TIME	day:1777-04-30	Gauss
TIME	day:1855-02-23	Gauss
TIME	century:17	Gauss
TIME	...other 7 terms	Gauss

KE4IR

Retrieval Model

- Inspired to the Vector Space Model (VSM)
- Queries and documents are represented as vector of terms

astronomers influenced by Gauss

$q = (q_i)$

$d = (d_i)$

$$sim = d \cdot q$$

- sim(d,q) > 0 ⇒ document is relevant for the query

KE4IR

Building the vectors

- Concatenation of layer-specific vectors

- Three ingredients:
  - Term Frequency (tf)
  - Inverse Document Frequency (idf)
  - Layer weight (w)

KE4IR

Building the vectors: example

astronomers influenced by Gauss

Layer	Term	Mentions
TEXTUAL	astronom	astronomers
TEXTUAL	influenc	influenced
TEXTUAL	gauss	Gauss
URI	dbpedia:Carl Friedrich Gauss	Gauss
TYPE	yago:GermanMathematicians	Gauss
TYPE	yago:NumberTheorists	Gauss
TYPE	yago:FellowsOfTheRoyalSociety	Gauss
TYPE	*...other 18 terms ...*	Gauss
TYPE	yago:Astronomer109818343	astronomers, Gauss
TYPE	yago:Physicist110428004	astronomers, Gauss
TYPE	yago:Person100007846	astronomers, Gauss
TYPE	*...other 9 terms ...*	astronomers, Gauss
FRAME	⟨Subjective influence-influence.v, Carl . . . Gauss⟩	influenced
FRAME	⟨Subjective influence, Carl Friedrich Gauss⟩	influenced
FRAME	⟨Frame, Carl Friedrich Gauss⟩	influenced
TIME	day:1777-04-30	Gauss
TIME	day:1855-02-23	Gauss
TIME	century:17	Gauss
TIME	...other 7 terms	Gauss

KE4IR

Building the vectors: example

astronomers influenced by Gauss

Layer	Term	Mentions	tf_i
TEXTUAL	astronom	astronomers	1.0
TEXTUAL	influenc	influenced	1.0
TEXTUAL	gauss	Gauss	1.0
URI	dbpedia:Carl Friedrich Gauss	Gauss	1.0
TYPE	yago:GermanMathematicians	Gauss	0.030
TYPE	yago:NumberTheorists	Gauss	0.030
TYPE	yago:FellowsOfTheRoyalSociety	Gauss	0.030
TYPE	*...other 18 terms ...*	Gauss	0.030
TYPE	yago:Astronomer109818343	astronomers, Gauss	0.114
TYPE	yago:Physicist110428004	astronomers, Gauss	0.114
TYPE	yago:Person100007846	astronomers, Gauss	0.114
TYPE	*...other 9 terms ...*	astronomers, Gauss	0.114
FRAME	⟨Subjective influence-influence.v, Carl . . . Gauss⟩	influenced	0.333
FRAME	⟨Subjective influence, Carl Friedrich Gauss⟩	influenced	0.333
FRAME	⟨Frame, Carl Friedrich Gauss⟩	influenced	0.333
TIME	day:1777-04-30	Gauss	0.1
TIME	day:1855-02-23	Gauss	0.1
TIME	century:17	Gauss	0.1
TIME	...other 7 terms	Gauss	0.1

KE4IR

Building the vectors: example

astronomers influenced by Gauss

Layer	Term	Mentions	tf_i	idf_i
TEXTUAL	astronom	astronomers	1.0	2.018
TEXTUAL	influenc	influenced	1.0	3.404
TEXTUAL	gauss	Gauss	1.0	1.568
URI	dbpedia:Carl Friedrich Gauss	Gauss	1.0	3.404
TYPE	yago:GermanMathematicians	Gauss	0.030	2.624
TYPE	yago:NumberTheorists	Gauss	0.030	2.583
TYPE	yago:FellowsOfTheRoyalSociety	Gauss	0.030	1.057
TYPE	*...other 18 terms ...*	Gauss	0.030	...
TYPE	yago:Astronomer109818343	astronomers, Gauss	0.114	1.432
TYPE	yago:Physicist110428004	astronomers, Gauss	0.114	0.958
TYPE	yago:Person100007846	astronomers, Gauss	0.114	0.003
TYPE	*...other 9 terms ...*	astronomers, Gauss	0.114	...
FRAME	⟨Subjective influence-influence.v, Carl . . . Gauss⟩	influenced	0.333	5.802
FRAME	⟨Subjective influence, Carl Friedrich Gauss⟩	influenced	0.333	5.802
FRAME	⟨Frame, Carl Friedrich Gauss⟩	influenced	0.333	3.499
TIME	day:1777-04-30	Gauss	0.1	3.404
TIME	day:1855-02-23	Gauss	0.1	3.404
TIME	century:17	Gauss	0.1	0.196
TIME	...other 7 terms	Gauss	0.1	...

KE4IR

Building the vectors: example

astronomers influenced by Gauss

Layer	Term	Mentions	tf_i	idf_i	w_i
TEXTUAL	astronom	astronomers	1.0	2.018	0.5
TEXTUAL	influenc	influenced	1.0	3.404	0.5
TEXTUAL	gauss	Gauss	1.0	1.568	0.5
URI	dbpedia:Carl Friedrich Gauss	Gauss	1.0	3.404	0.125
TYPE	yago:GermanMathematicians	Gauss	0.030	2.624	0.125
TYPE	yago:NumberTheorists	Gauss	0.030	2.583	0.125
TYPE	yago:FellowsOfTheRoyalSociety	Gauss	0.030	1.057	0.125
TYPE	*...other 18 terms ...*	Gauss	0.030	...	0.125
TYPE	yago:Astronomer109818343	astronomers, Gauss	0.114	1.432	0.125
TYPE	yago:Physicist110428004	astronomers, Gauss	0.114	0.958	0.125
TYPE	yago:Person100007846	astronomers, Gauss	0.114	0.003	0.125
TYPE	*...other 9 terms ...*	astronomers, Gauss	0.114	...	0.125
FRAME	⟨Subjective influence-influence.v, Carl . . . Gauss⟩	influenced	0.333	5.802	0.125
FRAME	⟨Subjective influence, Carl Friedrich Gauss⟩	influenced	0.333	5.802	0.125
FRAME	⟨Frame, Carl Friedrich Gauss⟩	influenced	0.333	3.499	0.125
TIME	day:1777-04-30	Gauss	0.1	3.404	0.125
TIME	day:1855-02-23	Gauss	0.1	3.404	0.125
TIME	century:17	Gauss	0.1	0.196	0.125
TIME	...other 7 terms	Gauss	0.1	...	0.125

KE4IR

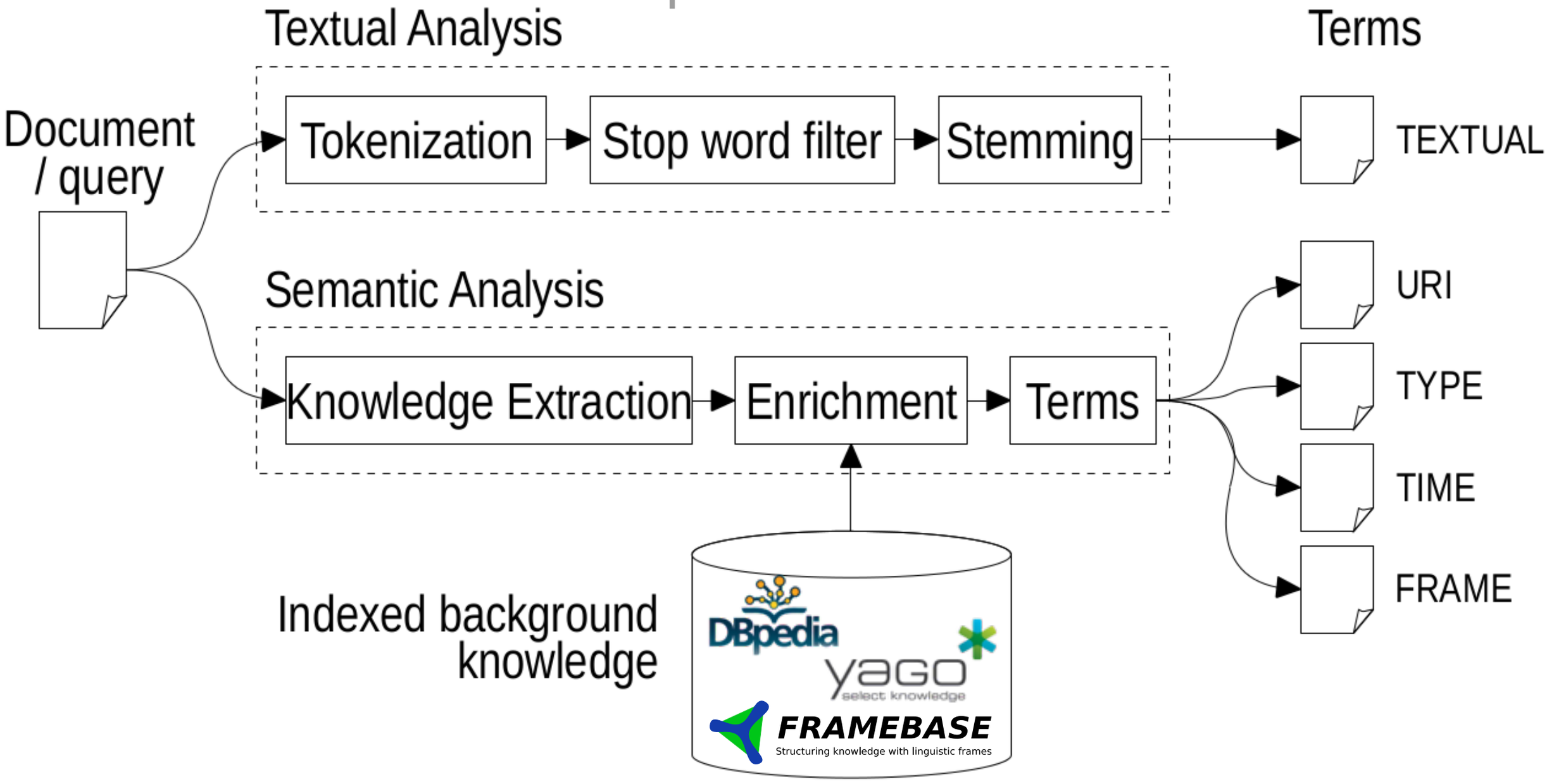
Building the vectors: example

astronomers influenced by Gauss

Layer	Term	Mentions	tf_i	idf_i	w_i	q_i
TEXTUAL	astronom	astronomers	1.0	2.018	0.5	1.009
TEXTUAL	influenc	influenced	1.0	3.404	0.5	1.702
TEXTUAL	gauss	Gauss	1.0	1.568	0.5	0.784
URI	dbpedia:Carl Friedrich Gauss	Gauss	1.0	3.404	0.125	0.426
TYPE	yago:GermanMathematicians	Gauss	0.030	2.624	0.125	0.010
TYPE	yago:NumberTheorists	Gauss	0.030	2.583	0.125	0.010
TYPE	yago:FellowsOfTheRoyalSociety	Gauss	0.030	1.057	0.125	0.004
TYPE	*...other 18 terms ...*	Gauss	0.030	...	0.125	...
TYPE	yago:Astronomer109818343	astronomers, Gauss	0.114	1.432	0.125	0.020
TYPE	yago:Physicist110428004	astronomers, Gauss	0.114	0.958	0.125	0.014
TYPE	yago:Person100007846	astronomers, Gauss	0.114	0.003	0.125	~0
TYPE	*...other 9 terms ...*	astronomers, Gauss	0.114	...	0.125	...
FRAME	⟨Subjective influence-influence.v, Carl . . . Gauss⟩	influenced	0.333	5.802	0.125	0.242
FRAME	⟨Subjective influence, Carl Friedrich Gauss⟩	influenced	0.333	5.802	0.125	0.242
FRAME	⟨Frame, Carl Friedrich Gauss⟩	influenced	0.333	3.499	0.125	0.146
TIME	day:1777-04-30	Gauss	0.1	3.404	0.125	0.043
TIME	day:1855-02-23	Gauss	0.1	3.404	0.125	0.043
TIME	century:17	Gauss	0.1	0.196	0.125	0.002
TIME	...other 7 terms	Gauss	0.1	...	0.125	...

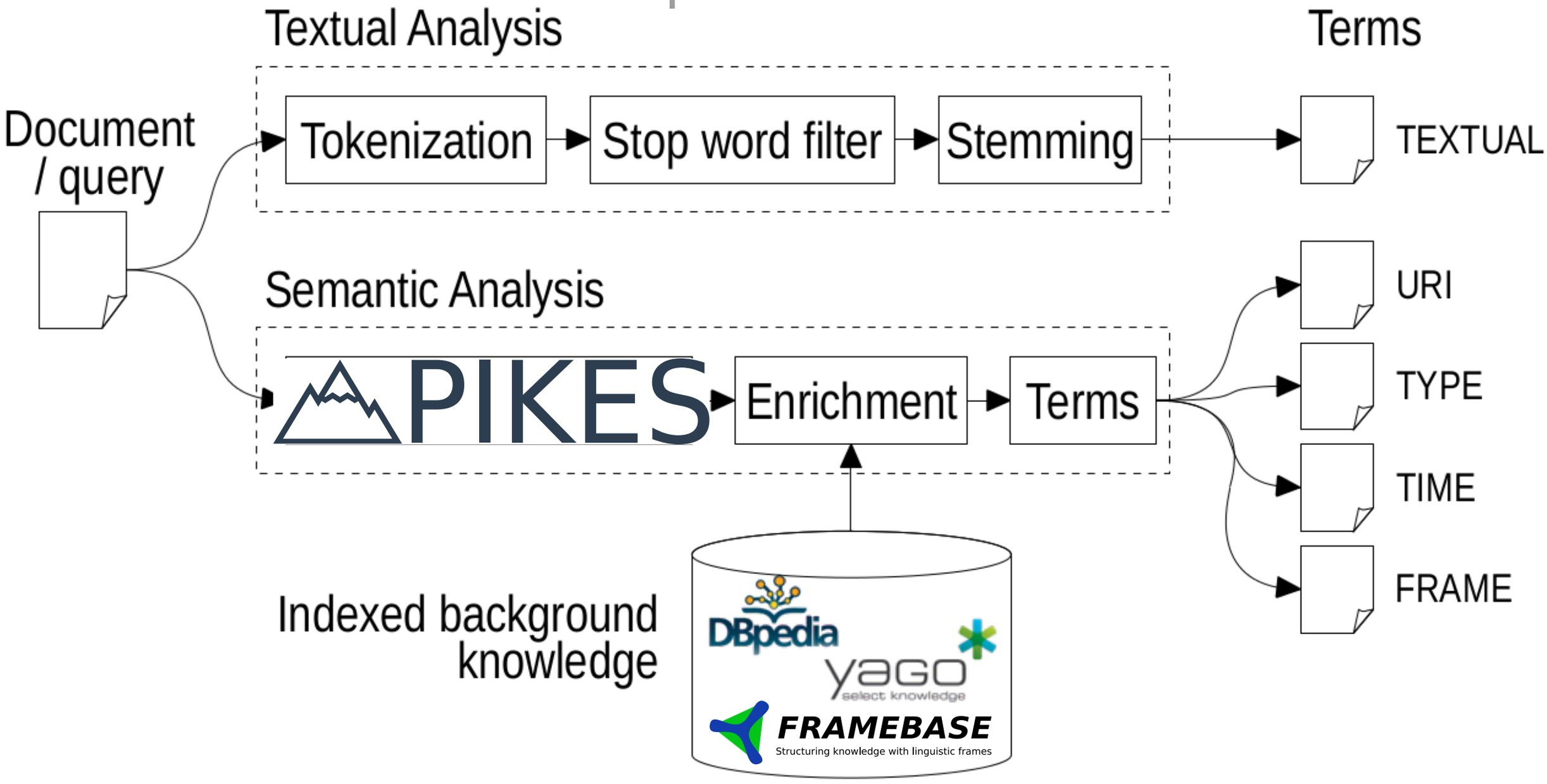
KE4IR

Implementation



KE4IR

Implementation



PIKES [ACM-SAC2016]

In a nutshell

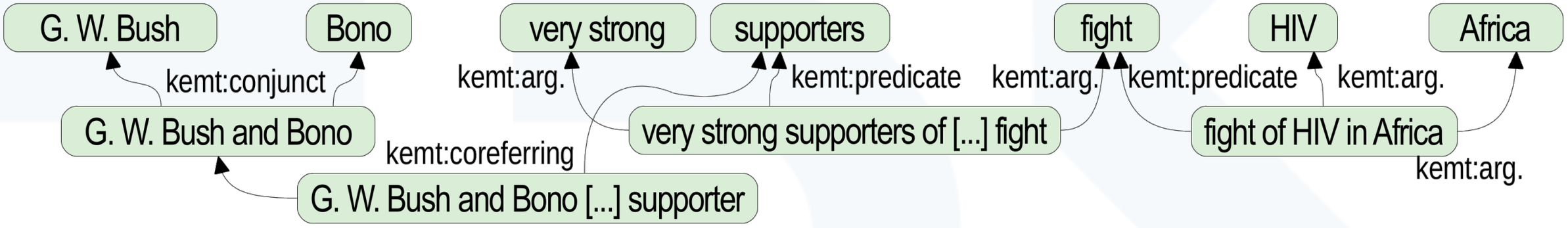
G. W. Bush and Bono are very strong supporters of the fight of HIV in Africa.

PIKES [ACM-SAC2016]

In a nutshell

G. W. Bush and Bono are very strong supporters of the fight of HIV in Africa.

Phase I: Linguistic Feature Extraction

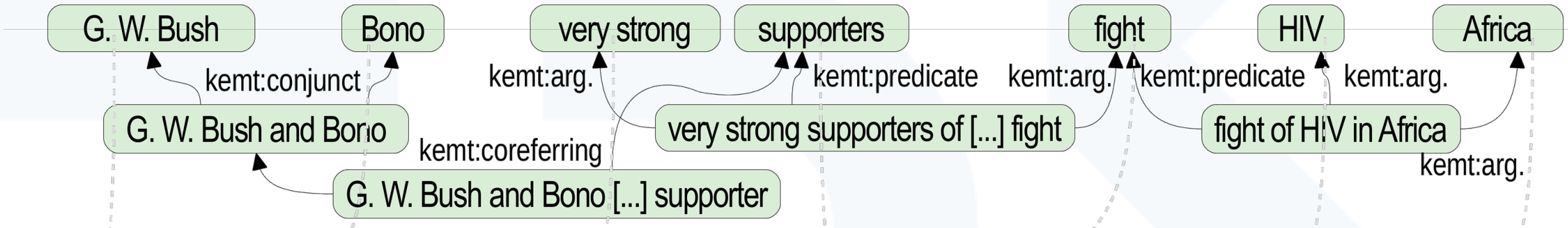


PIKES [ACM-SAC2016]

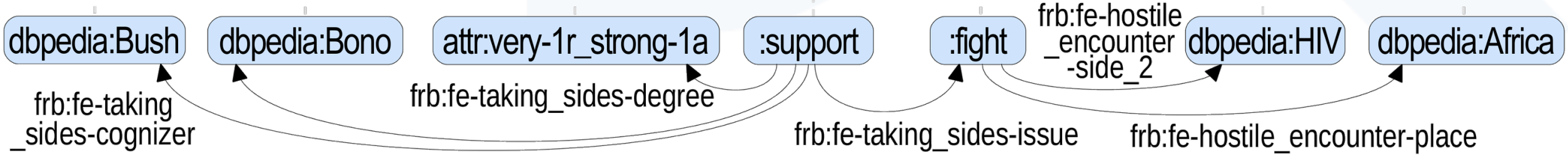
In a nutshell

G. W. Bush and Bono are very strong supporters of the fight of HIV in Africa.

Phase1: Linguistic Feature Extraction



Phase2: Knowledge Distillation



PIKES[ACM-SAC2016]

Main Characteristics

- State-of-the-art tool for frame-based ontology population
  - FrameBase Ontology Populator

- Modular nature

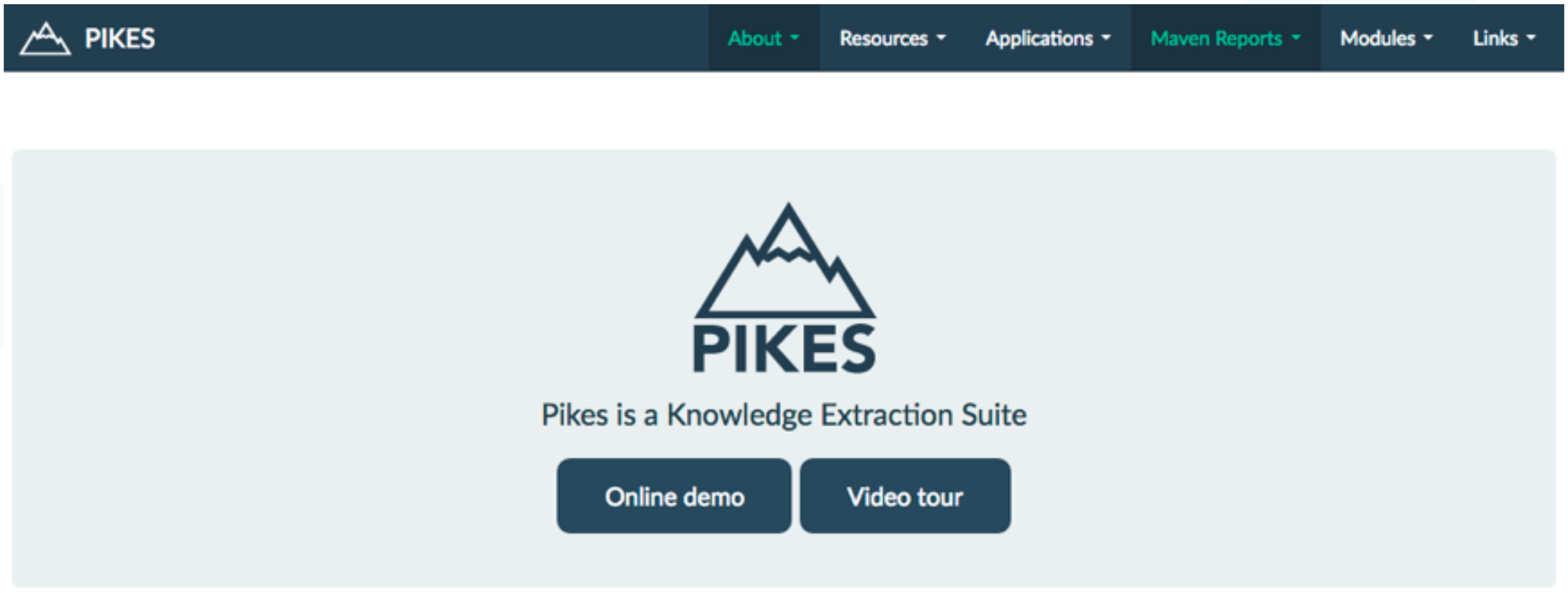
- All output exposed as RDF
  + Named Graph for knowledge tracing

- Efficiently process large corpora (700K tokens/hour)

PIKES [ACM-SAC2016]

Summary

http://pikes.fbk.eu/



About

PIKES is a Java-based suite that extracts knowledge from textual resources. The tool implements a rule-based strategy that reinterprets the output of semantic role labelling (SRL) tools in light of other linguistic analyses, such as dependency parsing or co-reference resolution, thus properly capturing and formalizing in RDF important linguistic aspects such as argument nominalization, frame-frame relations, and group entities.

Features

- Argument nominalization using SRL
- Frame-frame relations extractions
- Entity grouping exploiting linking and co-reference
- Extensible and replaceable NLP pipeline
- Interlinked three-layer representation model exposed as RDF
- Instance RDF triples annotated with detailed information of the mentions (via named graph)
- REST API service included, built on top of Grizzly
- Based on Java 8 and RDFpro

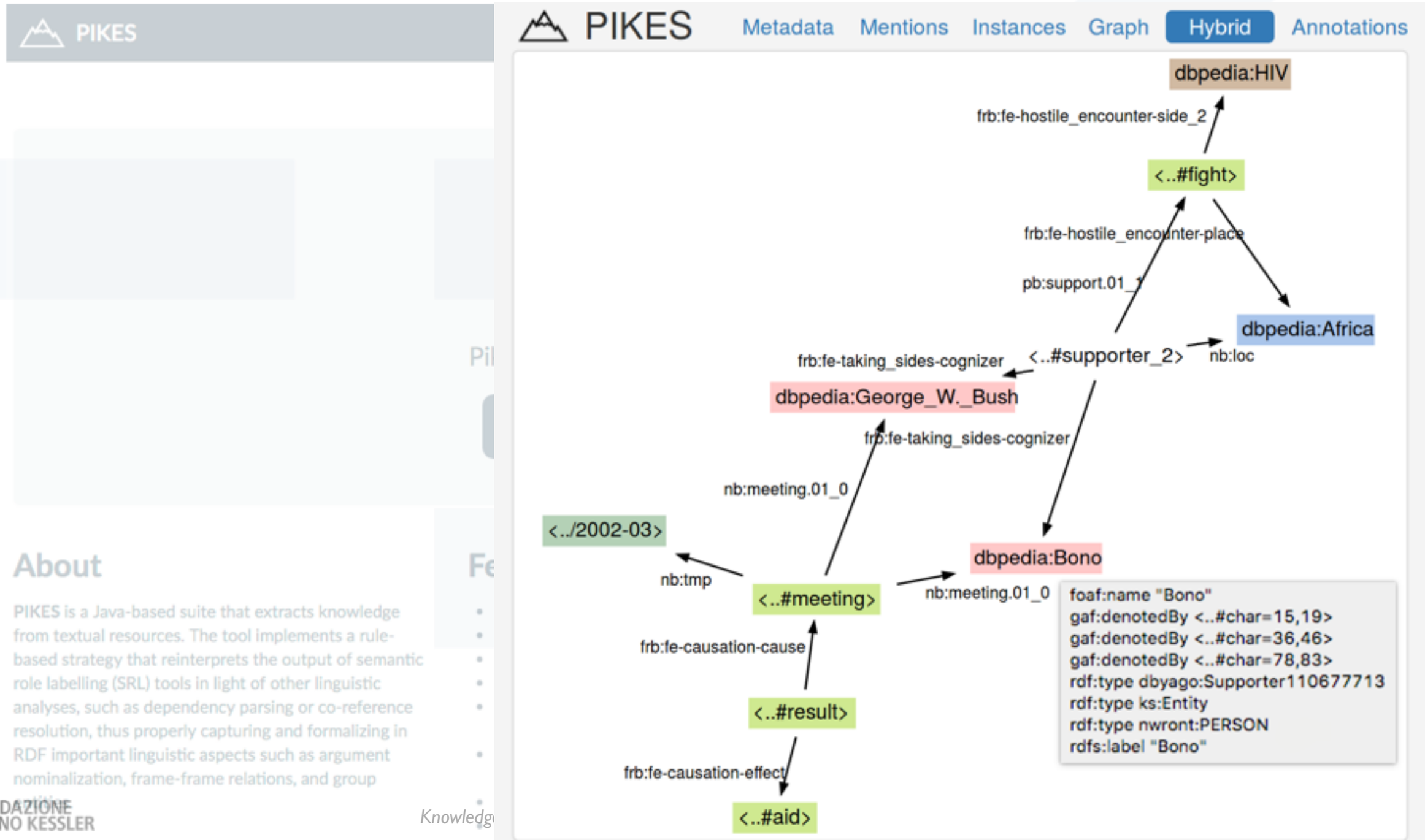
News

- 2016-03-18 Added Using PIKES for Information Retrieval section
- 2016-02-23 Paper on using PIKES for Information Retrieval accepted at ESWC 2016 conference
- 2015-11-24 Paper accepted at SAC 2016 conference, SWA track
- 2015-08-07 Demo accepted at ISWC 2015 conference
- 2015-07-08 Restyling of the Try-it-out demo
- 2015-07-06 New Demo Video!

PIKES [ACM-SAC2016]

Summary

http://pikes.fbk.eu/



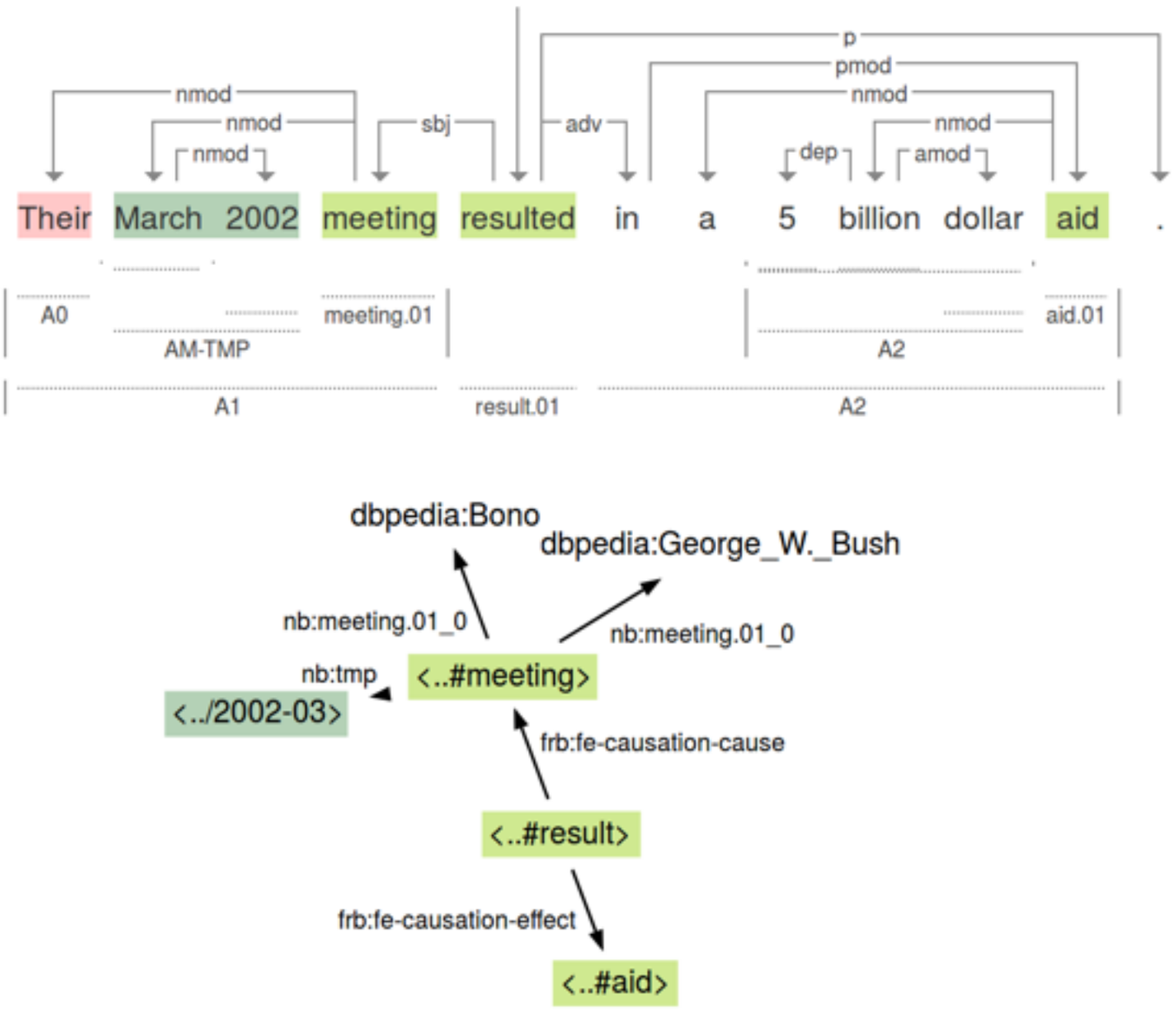
PIKES [ACM-SAC2016]

http://pikes.fbk.eu/

PIKES Metadata Mentions Instances Graph Hybrid Annotations

S1 G. W. Bush and Bono are very strong supporters of the fight of HIV in Africa.

S2



PIKES [ACM-SAC2016]

Summary

http://pikes.fbk.eu/

PIKES Metadata Mentions Instances Graph Hybrid Annotations

S1 G. W. Bush and Bono are very strong supporters of the fight of HIV in Africa.

S2

PIKES Metadata Mentions Instances Graph Hybrid Annotations

Showing 1 to 11 of 11 entries (filtered from 31 total entries) Search: dbpedia

id	nif:anchorOf	rdf:type	mention attributes	ks:denotes/ks:implies	ks:expresses
<../char=63,66>	HIV	ks:EntityMention ks:NameMention ks:Mention		dbpedia:HIV	1. dbpedia:HIV, foaf:name, HIV 2. dbpedia:HIV, rdf:type, ks:Entity 3. dbpedia:HIV, rdfs:label, HIV
<../char=0,10>	G. W. Bush	ks:NameMention ks:Mention ks:EntityMention		dbpedia:George_W._Bush	1. dbpedia:George_W._Bush, rdf:type, ks:Entity 2. dbpedia:George_W._Bush, foaf:name, G. W. Bush 3. dbpedia:George_W._Bush, rdf:type, nwront:PERSON
<../char=78,83>	Their	ks:Mention ks:EntityMention		dbpedia:Bono dbpedia:George_W._Bush	1. dbpedia:Bono, rdf:type, ks:Entity 2. dbpedia:George_W._Bush, rdf:type, ks:Entity
<../char=36,46>	supporters	ks:PredicateMention ks:ParticipationMention ks:EntityMention ks:Mention	ks:lemma: supporter ks:plural: true ks:sst: <../noun.person> ks:synset: <../10677713-n>	<../supporter_2> dbpedia:Bono dbpedia:George_W._Bush	1. <../supporter_2>, rdf:type, fs:frame-Taking_sides-back.v 2. <../supporter_2>, rdf:type, ks:Entity 3. <../supporter_2>, rdf:type, sem:Event

KE4IR

Evaluation Setup

- 331 documents, 35 queries [Waitelonis et al, 2015]

- Multi-value relevances (1=irrelevant, 5=relevant)

- Diverse queries: from keyword-base search to queries requiring semantic capabilities

KE4IR

Evaluation Setup

- 2 baselines:
  - Google custom search API
  - Textual layer only (~Lucene)

- Measures: $Prec_{1,5,10}$, MAP, MAP_{10}, NDCG, $NDCG_{10}$

- Same weights for textual and semantic layers:
  - TEXTUAL (50%)
  - URI (12,5%), TYPE (12,5%), FRAME (12,5%), TIME (12,5%)

KE4IR

Evaluation Results: Comparison with the baselines

Approach/System	$Prec_1$	$Prec_5$	$Prec_{10}$	NDCG	$NDCG_{10}$	MAP	MAP_{10}
Google	0.543	0.411	0.343	0.434	0.405	0.255	0.219
Textual	0.943	0.669	0.453	0.832	0.782	0.733	0.681
KE4IR	0.971	0.680	0.474	0.854	0.806	0.758	0.713

KE4IR

Evaluation Results: Comparison with the baselines

Approach/System	$Prec_1$	$Prec_5$	$Prec_{10}$	NDCG	$NDCG_{10}$	MAP	MAP_{10}
Google	0.543	0.411	0.343	0.434	0.405	0.255	0.219
Textual	0.943	0.669	0.453	0.832	0.782	0.733	0.681
KE4IR	0.971	0.680	0.474	0.854	0.806	0.758	0.713
KE4IR vs. Textual	3.03%	1.71%	4.55%	2.64%	2.99%	3.50%	4.74%

KE4IR

Evaluation Results: Comparison with the baselines

Approach/System	$Prec_1$	$Prec_5$	$Prec_{10}$	NDCG	$NDCG_{10}$	MAP	MAP_{10}
Google	0.543	0.411	0.343	0.434	0.405	0.255	0.219
Textual	0.943	0.669	0.453	0.832	0.782	0.733	0.681
KE4IR	0.971	0.680	0.474	0.854	0.806	0.758	0.713
KE4IR vs. Textual	3.03%	1.71%	4.55%	2.64%	2.99%	3.50%	4.74%

statistically significant

KE4IR

Evaluation Results: Comparison with the baselines

Approach/System	$Prec_1$	$Prec_5$	$Prec_{10}$	NDCG	$NDCG_{10}$	MAP	MAP_{10}
Google	0.543	0.411	0.343	0.434	0.405	0.255	0.219
Textual	0.943	0.669	0.453	0.832	0.782	0.733	0.681
KE4IR	0.971	0.680	0.474	0.854	0.806	0.758	0.713
KE4IR vs. Textual	3.03%	1.71%	4.55%	2.64%	2.99%	3.50%	4.74%

statistically significant

Knowledge Extraction positively affects the Document Retrieval performances!

KE4IR

Evaluation Results: Impact of various layer combinations

Layers (TEXTUAL+)	$Prec_1$	$Prec_5$	$Prec_{10}$	NDCG	$NDCG_{10}$	MAP	MAP_{10}
URI,TYPE,FRAME,TIME	0.971	0.680	0.474	0.854	0.806	0.758	0.713
URI,TYPE,FRAME	0.971	0.680	0.474	0.853	0.804	0.757	0.712
URI,TYPE,TIME	0.971	0.680	0.474	0.851	0.802	0.757	0.712
URI,TYPE	0.971	0.680	0.474	0.849	0.801	0.755	0.710
URI,FRAME,TIME	0.971	0.674	0.465	0.844	0.796	0.750	0.702
URI,FRAME	0.971	0.674	0.465	0.842	0.795	0.749	0.702
URI,TIME	0.971	0.674	0.465	0.840	0.791	0.747	0.700
URI	0.971	0.674	0.465	0.837	0.791	0.747	0.700
TYPE,FRAME,TIME	0.943	0.674	0.471	0.848	0.799	0.745	0.700
TYPE,TIME	0.943	0.674	0.471	0.843	0.794	0.743	0.697
TYPE,FRAME	0.943	0.674	0.468	0.847	0.797	0.743	0.695
FRAME,TIME	0.943	0.674	0.462	0.842	0.793	0.741	0.693
TYPE	0.943	0.674	0.468	0.842	0.792	0.740	0.693
TIME	0.943	0.669	0.462	0.836	0.786	0.737	0.689
FRAME	0.943	0.674	0.453	0.839	0.789	0.737	0.686
(only textual)	0.943	0.669	0.453	0.832	0.782	0.733	0.681

KE4IR

Evaluation Results: Impact of various layer combinations

Layers (TEXTUAL+)	$Prec_1$	$Prec_5$	$Prec_{10}$	NDCG	$NDCG_{10}$	MAP	MAP_{10}
URI,TYPE,FRAME,TIME	0.971	0.680	0.474	0.854	0.806	0.758	0.713
URI,TYPE,FRAME	0.971	0.680	0.474	0.853	0.804	0.757	0.712
URI,TYPE,TIME	0.971	0.680	0.474	0.851	0.802	0.757	0.712
URI,TYPE	0.971	0.680	0.474	0.849	0.801	0.755	0.710
URI,FRAME,TIME	0.971	0.674	0.465	0.844	0.796	0.750	0.702
URI,FRAME	0.971	0.674	0.465	0.842	0.795	0.749	0.702
URI,TIME	0.971	0.674	0.465	0.840	0.791	0.747	0.700
URI	0.971	0.674	0.465	0.837	0.791	0.747	0.700
TYPE,FRAME,TIME	0.943	0.674	0.471	0.848	0.799	0.745	0.700
TYPE,TIME	0.943	0.674	0.471	0.843	0.794	0.743	0.697
TYPE,FRAME	0.943	0.674	0.468	0.847	0.797	0.743	0.695
FRAME,TIME	0.943	0.674	0.462	0.842	0.793	0.741	0.693
TYPE	0.943	0.674	0.468	0.842	0.792	0.740	0.693
TIME	0.943	0.669	0.462	0.836	0.786	0.737	0.689
FRAME	0.943	0.674	0.453	0.839	0.789	0.737	0.686
(only textual)	0.943	0.669	0.453	0.832	0.782	0.733	0.681

KE4IR

Evaluation Results: Query-by-query analysis

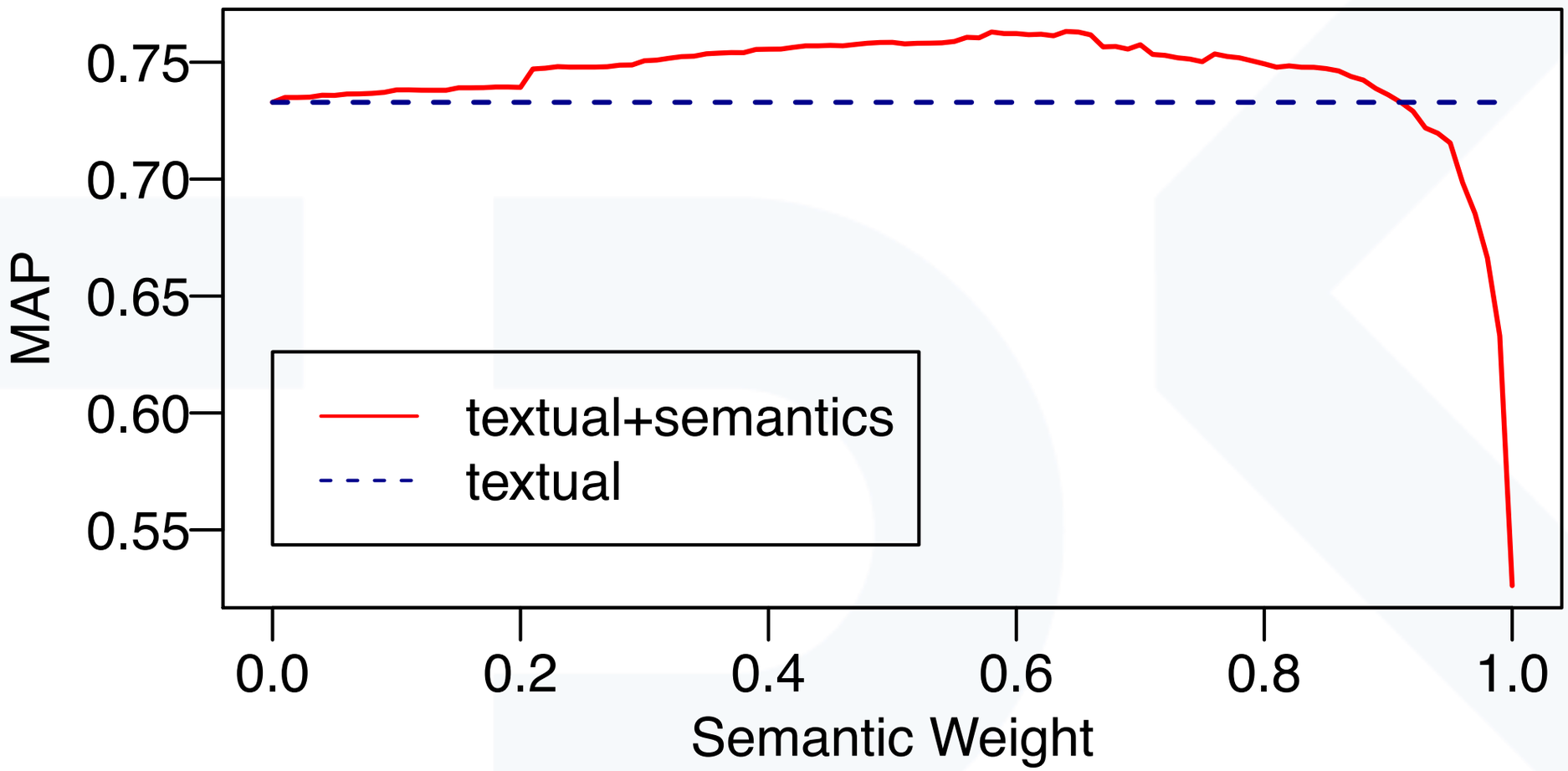
- General Remarks
  - TYPE & URI: more frequent, less "reliable"
  - FRAME & TIME: less frequent, positively impact

- Analysis on selected examples

Query Text	Δ NDCG@10	Δ MAP
Nazis confiscate or destroy art and literature	0.154	0.099
Modern Age in English Literature	-0.117	-0.095
Napoleon's Russian Campaign	0.151	0.147
First woman who won a Nobel Prize	0	0

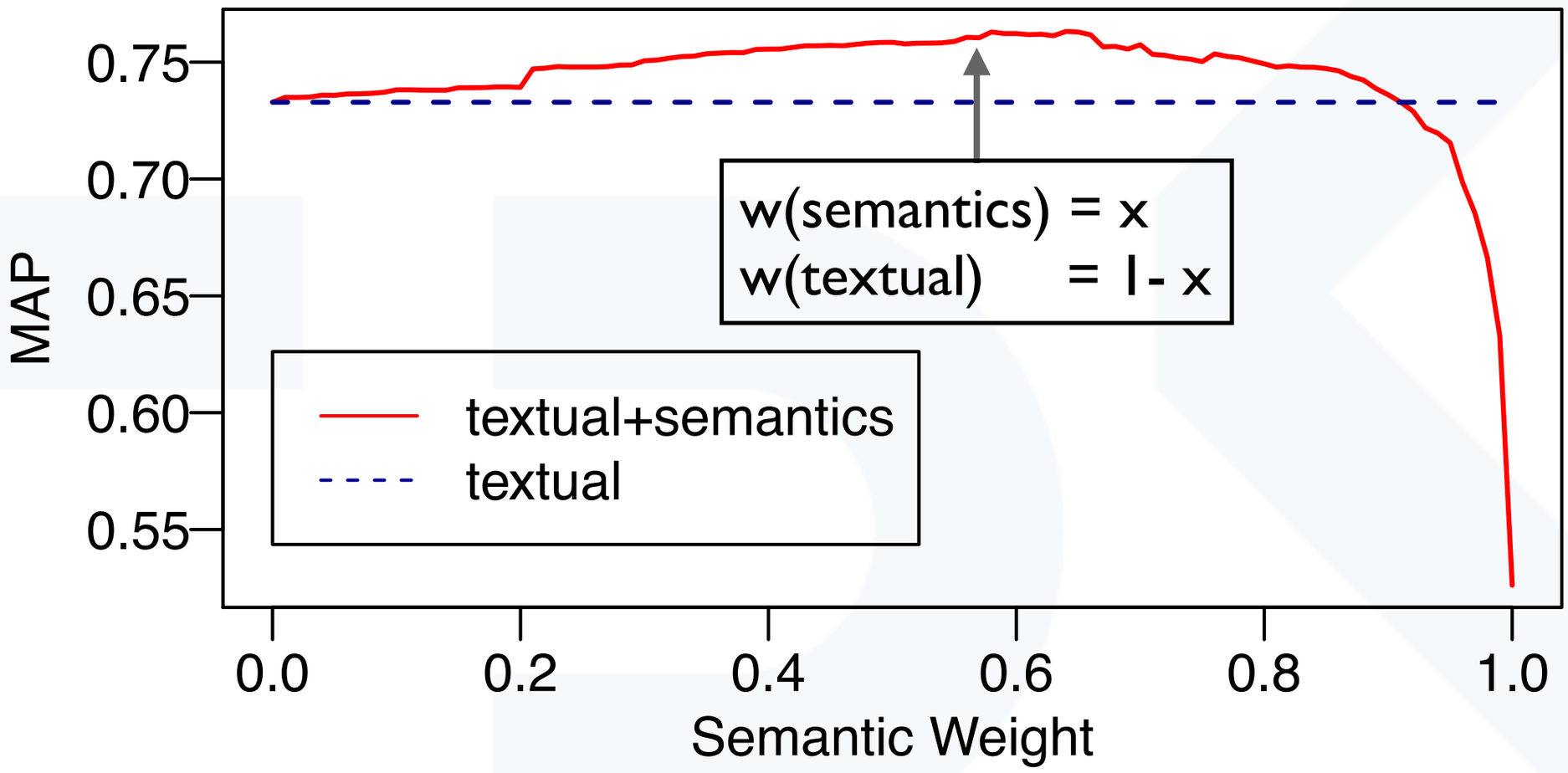
KE4IR

Balancing Textual vs Semantic Content



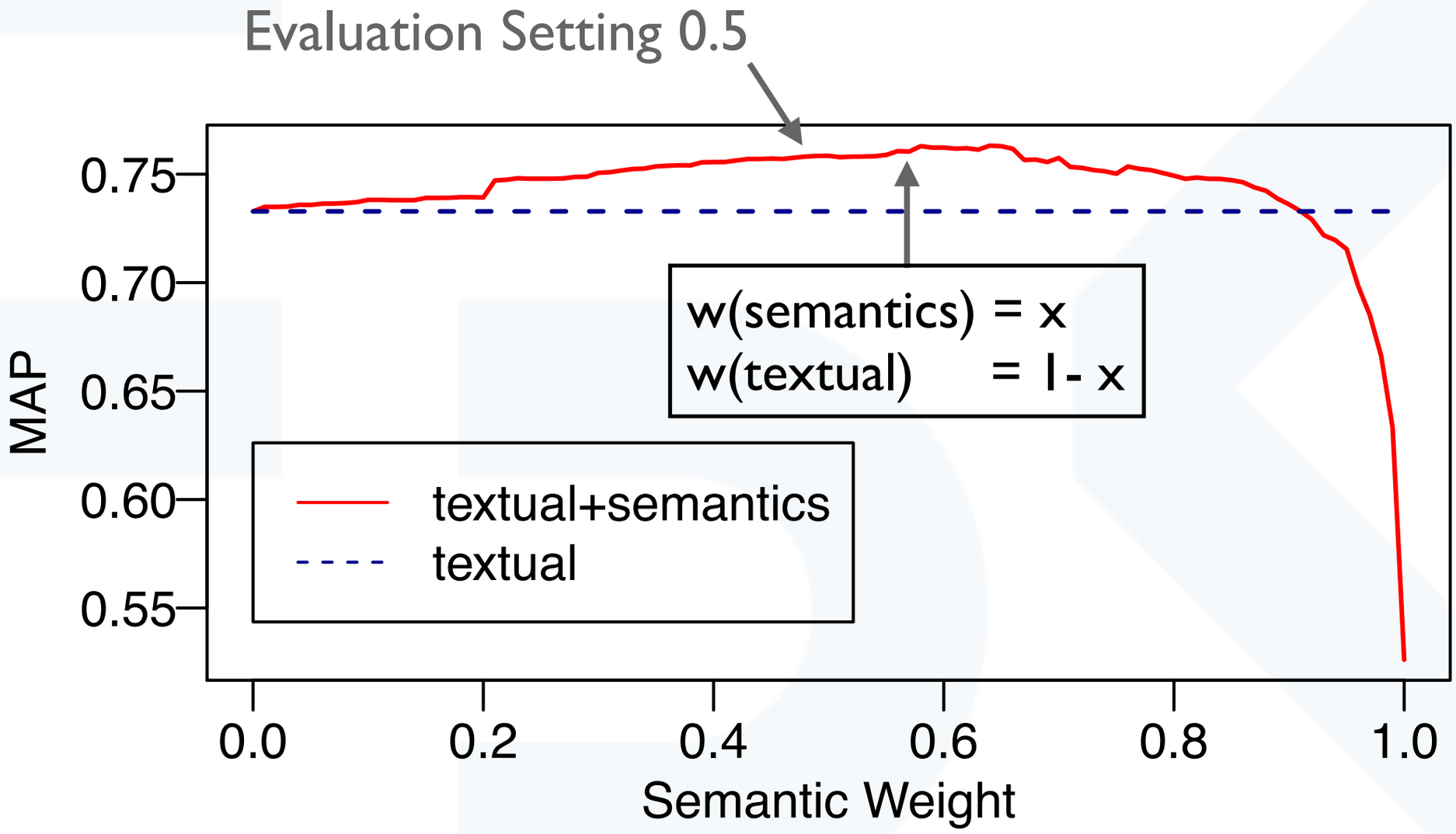
KE4IR

Balancing Textual vs Semantic Content



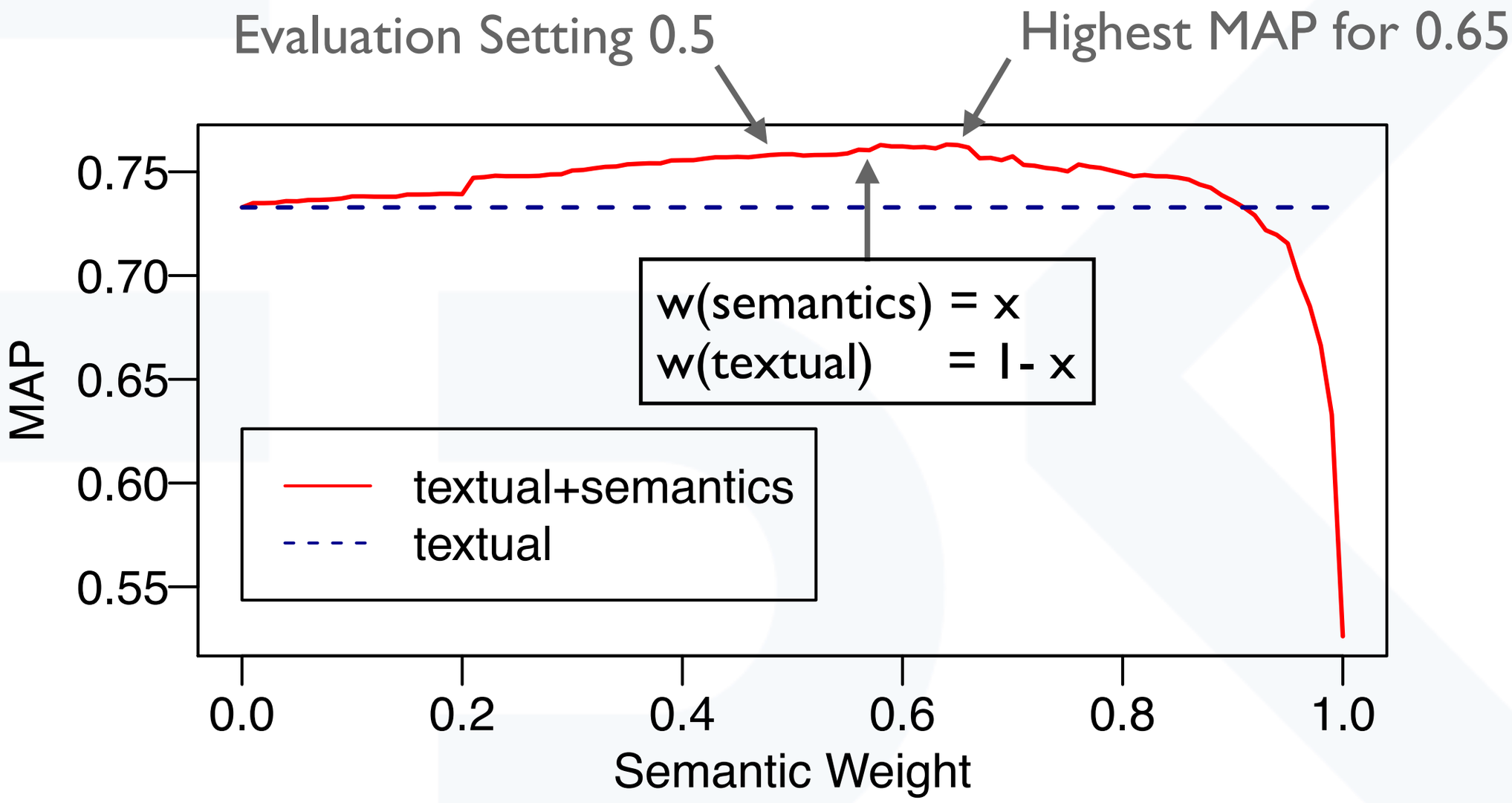
KE4IR

Balancing Textual vs Semantic Content



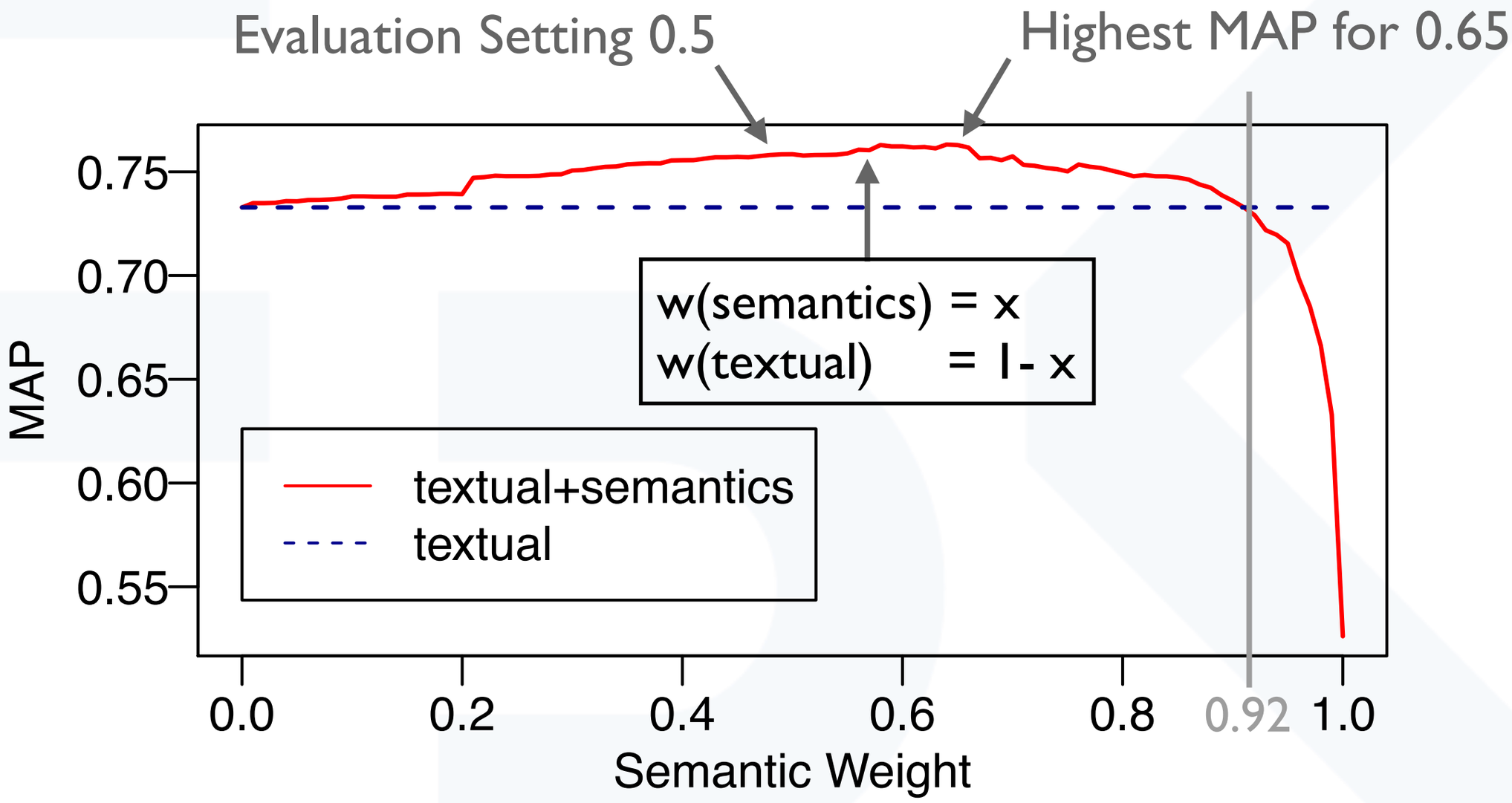
KE4IR

Balancing Textual vs Semantic Content



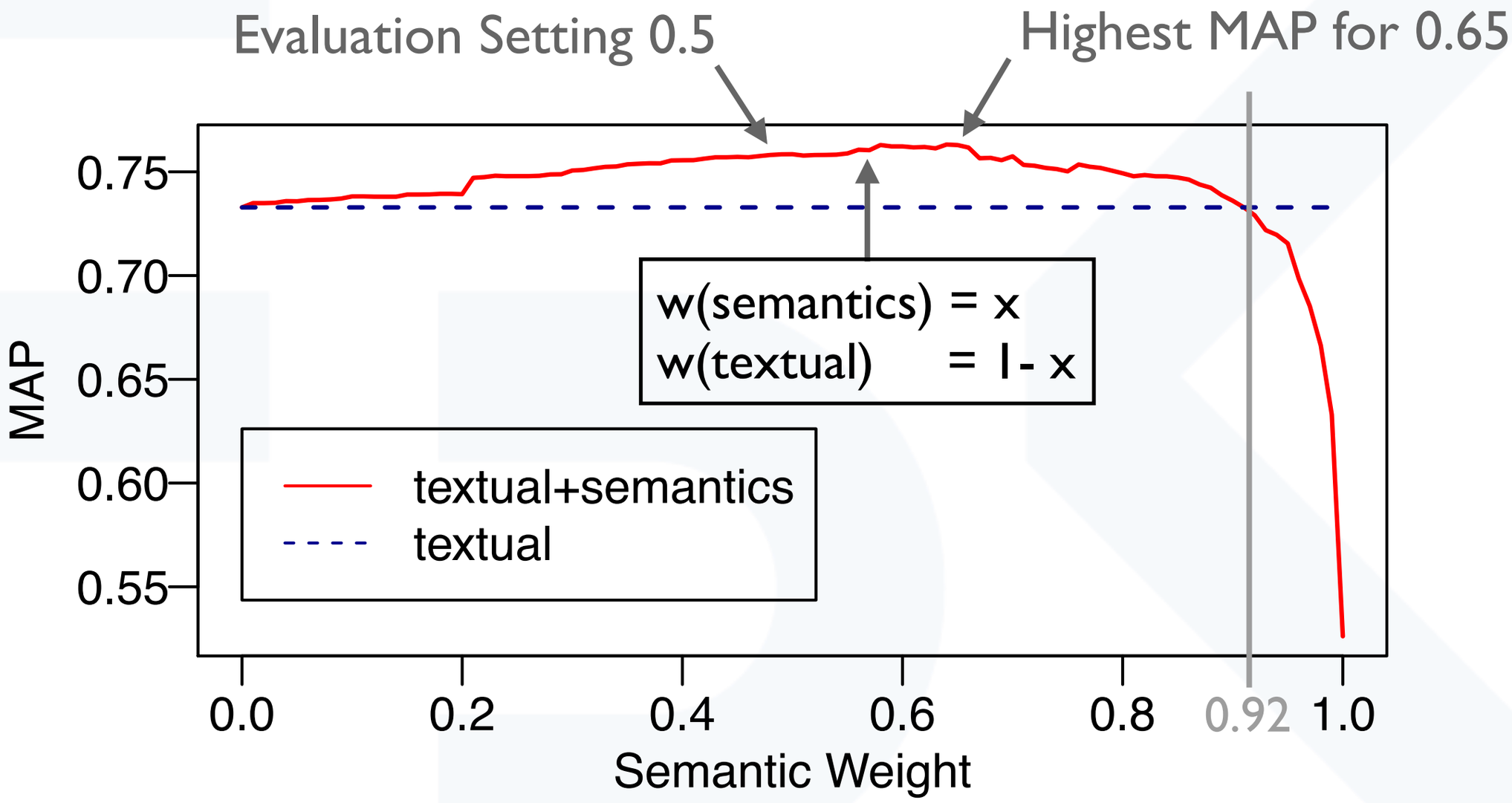
KE4IR

Balancing Textual vs Semantic Content



KE4IR

Balancing Textual vs Semantic Content

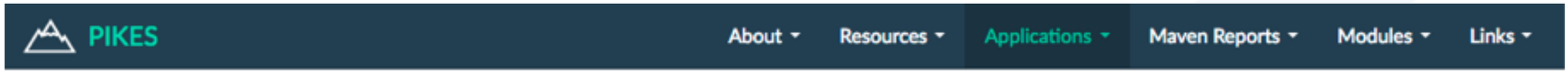


Too Much Semantics Will Kill You!

KE4IR

Evaluation Material

http://pikes.fbk.eu/ke4ir/



Using PIKES for Information Retrieval

This page provides additional details on the use of Knowledge Extraction techniques, as implemented in PIKES, to improve the performances of Information Retrieval. The proposed approach is fully implemented in the following paper:

- **Knowledge Extraction for Information Retrieval**
  By Francesco Corcoglioniti, Mauro Dragoni, Marco Rospocher, and Alessio Palmero Aprosio.
  In Proceedings of the 13th European Semantic Web Conference (ESWC2016), Anissaras, Crete, Greece, May 29-June 2, 2016
  [bib] [pre-print/mirror]

Here, we provide a brief overview of the approach, make available for downlaod all the code and data used in the evaluation (to allow reproducibility of our results), and provide all the reports and the additional material we produced as part of the evaluation.

Approach

The goal in Information Retrieval is to determine, for a given text query, the relevant documents in a text collection, ranking them according to their relevance degree for the query.

In our approach, named **KE4IR** (read: *kee-fer*) and implemented on top of PIKES and Apache Lucene, both queries and documents are processed to extract semantic terms pertaining to the following semantic layers:

- **URI layer**, containing URIs of entities mentioned in the text, disambiguated against DBpedia;
- **TYPE layer**, containing YAGO classes associated to noun phrases in the text, extracted via disambiguation to WordNet (plus WordNet to YAGO mappings) or imported from DBpedia after entity linking;
- **TIME layer**, containing temporal values explicitly expressed in the text, or imported from DBpedia after entity linking;
- **FRAME layer**, containing <frame type, participant> pairs where the frame results from knowledge extraction and the participant is a disambiguated entity.

Conclusions

- Exploiting the knowledge extracted from queries and documents improves IR performances

- Evaluation results legitimise testing KE4IR in real-world situations

- Looking ahead to the future….
  - larger document collections (e.g., TREC WT10g, ClueWeb)
  - favouring precision over recall in KE?
  - domain-adaptation

KE4IR

pikes.fbk.eu/ke4ir

PREMON

premon.fbk.eu

PIKES

pikes.fbk.eu

Marco Rospocher

FONDAZIONE BRUNO KESSLER

rospocher@fbk.eu

dkm.fbk.eu/rospocher

@marcorospocher

RDF pro

rdfpro.fbk.eu

KnowledgeStore

knowledgestore.fbk.eu

MoKi

the Modelling WiKi ---

moki.fbk.eu