Argumentation Technology for Artificial Intelligence

Part 3: Argument Retrieval

Philipp Cimiano, **Khalid Al-Khatib** / **Benno Stein**, Henning Wachsmuth

September 21th, 2020

UNIVERSITÄT BIELEFELD

Bauhaus-Universität Weimar

PADERBORN UNIVERSITY

Basic Argument Model

Conclusion	*We should not colonize the Moon.*
Premise 1	*Colonizing Moon is just about funding for NASA.*
Premise 2	*Moon's gravity is too low for human health.*
Premise 3	*Human survival demands fighting global warming, not Moonbase.*

Basic Argument Model

Conclusion	*We should not colonize the Moon.*
Premise 1	*Colonizing Moon is just about funding for NASA.*
Premise 2	*Moon's gravity is too low for human health.*
Premise 3	*Human survival demands fighting global warming, not Moonbase.*

Argument:

- A conclusion supported by premises. [Walton et al. 2008]
  - Conclusion and premises are considered as propositions.
  - Assignment of truth values to the propositions:

  $\mathcal{I}($"*Colonizing Moon is just about funding for NASA*"$) = 1,\ \mathcal{I}($"*Moon's gravity ...*"$) = 1,\ \ldots$

- Conveys a stance on a controversial topic. [Freeley and Steinberg, 2009]

- The mechanism to draw the conclusion from the premises is informal.
  - Implicit premises (Enthymemes)

Basic Argument Model

Conclusion	*We should not colonize the Moon.*
Premise 1	*Colonizing Moon is just about funding for NASA.*
Premise 2	*Moon's gravity is too low for human health.*
Premise 3	*Human survival demands fighting global warming, not Moonbase.*

Argument:

- A conclusion supported by premises. [Walton et al. 2008]
  - Conclusion and premises are considered as propositions.
  - Assignment of truth values to the propositions:
  $\mathcal{I}($"*Colonizing Moon is just about funding for NASA*"$) = 1,\ \mathcal{I}($"*Moon's gravity ...*"$) = 1,\ \ldots$

- Conveys a stance on a controversial topic. [Freeley and Steinberg, 2009]

- The mechanism to draw the conclusion from the premises is informal.
  - Implicit premises (Enthymemes)

Basic Argument Model

Conclusion	*We should not colonize the Moon.*
Premise 1	*Colonizing Moon is just about funding for NASA.*
Premise 2	*Moon's gravity is too low for human health.*
Premise 3	*Human survival demands fighting global warming, not Moonbase.*

Argument:

- A conclusion supported by premises. [Walton et al. 2008]
  - Conclusion and premises are considered as propositions.
  - Assignment of truth values to the propositions:
  $\mathcal{I}($*"Colonizing Moon is just about funding for NASA"*$) = 1, \ \mathcal{I}($*"Moon's gravity ..."*$) = 1, \ \ldots$
- Conveys a stance on a controversial topic. [Freeley and Steinberg, 2009]
- The mechanism to draw the conclusion from the premises is informal.
  - Implicit premises (Enthymemes)

Thesis / Major claim t *Space colonization is not needed for human survival.*

A_{pro}

c_1 *We should not colonize the Moon.*

P_1

p_1 *Colonizing Moon is just about funding for NASA.*

p_2 *Moon's gravity is too low for human health.*

p_3 *Human survival demands fighting global warming, not Moonbase.*

Thesis / Major claim t *Space colonization is not needed for human survival.*

c_1 *We should not colonize the Moon.*

A_{pro} P_1

p_1 *Colonizing Moon is just about funding for NASA.*

p_2 *Moon's gravity is too low for human health.*

p_3 *Human survival demands fighting global warming, not Moonbase.*

Note: $c_1 \succ t$

- "c_1 supports t" (entailment in a cogent, nonobligatory sense)
- "t is compatible with c_1" (but the real argumentation focus)

Thesis / Major claim t *Space colonization is not needed for human survival.*

A_{pro}

c_1 *We should not colonize the Moon.*

P_1

p_1 *Colonizing Moon is just about funding for NASA.*

p_2 *Moon's gravity is too low for human health.*

p_3 *Human survival demands fighting global warming, not Moonbase.*

A_{con}

c_2 *We should colonize the Moon.*

P_2

p_4 *Moon colony could be safeguard against asteroids.*

p_5 *Artificial gravity can overcome health risks on Moonbase.*

Thesis / Major claim t *Space colonization is not needed for human survival.*

A_{pro}

c_1 *We should not colonize the Moon.*

P_1

p_1 *Colonizing Moon is just about funding for NASA.*

p_2 *Moon's gravity is too low for human health.*

p_3 *Human survival demands fighting global warming, not Moonbase.*

A_{con}

c_2 *We should colonize the Moon.*

P_2

p_4 *Moon colony could be safeguard against asteroids.*

p_5 *Artificial gravity can overcome health risks on Moonbase.*

Note:

- $c_1 \approx \neg\, c_2$ "$\neg\, c_2$ is a paraphrase of c_1"

⇒ c_2 can be expressed as c_1 with opposite truth assignment, $\mathcal{I}(c_1) = 0$, $\mathcal{I}(c_2) = 1$

Argument Retrieval Problems

(1) Argument Relevance Π_{rel}

Query *Should we colonize the Moon?*

Argument Retrieval Problems

(1) Argument Relevance Π_{rel}

Query *Should we colonize the Moon?*

A_{pro}

c_1 *We should not colonize the Moon.*

P_1

p_1 *Colonizing Moon is just about funding for NASA.*

p_2 *Moon's gravity is too low for human health.*

p_3 *Human survival demands fighting global warming, not Moonbase.*

Given in Π_{rel} :

- information need, expressed as query, $q \in Q$
- set of arguments, $\mathbf{A} = \{(c_1, P_1), (c_2, P_2), \ldots, (c_n, P_n)\}$
- $*$ (possibly hidden) human selection of the relevant arguments, $\mathbf{A}^*_q$, $q \in Q$

Argument Retrieval Problems

(1) Argument Relevance Π_{rel}

Query *Should we colonize the Moon?*

A_{pro}

c_1 *We should not colonize the Moon.*

P_1:
- p_1 *Colonizing Moon is just about funding for NASA.*
- p_2 *Moon's gravity is too low for human health.*
- p_3 *Human survival demands fighting global warming, not Moonbase.*

Given in Π_{rel} :

- information need, expressed as query, $q \in Q$
- set of arguments, $\mathbf{A} = \{(c_1, P_1), (c_2, P_2), \ldots, (c_n, P_n)\}$
- $*$ (possibly hidden) human selection of the relevant arguments, $\mathbf{A}_q^*$, $q \in Q$

Sought in Π_{rel} :

- a relevance function $\rho : Q \times \mathbf{A} \rightarrow \{0, 1\}$, such that . . .

  the macro-averaged F-measure (precision, recall) regarding $\mathbf{A}_q^*$, $q \in Q$, is maximum

Argument Retrieval Problems

(2) Argument Ranking Π_{rank}

Query *Should we colonize the Moon?*

A_{pro}

c_1 *We should not colonize the Moon.*

P_1

p_1 *Colonizing Moon is just about funding for NASA.*

p_2 *Moon's gravity is too low for human health.*

p_3 *Human survival demands fighting global warming, not Moonbase.*

Given in Π_{rank} :

- information need, expressed as query, $q \in Q$
- set of relevant arguments, $\mathbf{A}_q = \{(c_1, P_1), (c_2, P_2), \ldots, (c_m, P_m)\}$
- (possibly hidden) human ranking of the relevant arguments, $\pi^*_{\mathbf{A}_q}$, $q \in Q$

Argument Retrieval Problems

(2) Argument Ranking Π_{rank}

Query *Should we colonize the Moon?*

$$A_{\text{pro}} \left\{ \begin{array}{ll} & c_1 \quad \textit{We should not colonize the Moon.} \\ \hline P_1 \left\{ \begin{array}{l} p_1 \\ p_2 \\ p_3 \end{array} \right. & \begin{array}{l} \textit{Colonizing Moon is just about funding for NASA.} \\ \textit{Moon's gravity is too low for human health.} \\ \textit{Human survival demands fighting global warming, not Moonbase.} \end{array} \end{array} \right.$$

Given in Π_{rank} :

- information need, expressed as query, $q \in Q$
- set of relevant arguments, $\mathbf{A}_q = \{(c_1, P_1), (c_2, P_2), \ldots, (c_m, P_m)\}$
- $*$ (possibly hidden) human ranking of the relevant arguments, $\pi^*_{\mathbf{A}_q}$, $q \in Q$

Sought in Π_{rank} :

- a ranking function $\sigma : Q \times \mathcal{P}(\mathbf{A}) \to \Pi$, such that ...

  the mean rank correlation $\overline{\tau}$ regarding $\pi^*_{\mathbf{A}_q}$, $q \in Q$, is maximum

Argument Retrieval Problems

(3) – (7) Further Problems

3. Π_{counter} Retrieve the "best" counterargument

   Given: query q, argument set $\mathbf{A}$, argument A

4. Π_{sameside} Retrieve (all) arguments with the same stance

   Given: argument set $\mathbf{A}$, argument A

5. Π_{argdoc} Is the document argumentative?

   Given: document d

6. Π_{argquery} Is the query argumentative?

   Given: query q

7. Π_{argsum} Summarize an argument.

   Given: argument A

Argument Retrieval Problems

(3) – (7) Further Problems

3. Π_{counter} Retrieve the "best" counterargument

   Given: query q, argument set $\mathbf{A}$, argument A

4. Π_{sameside} Retrieve (all) arguments with the same stance

   Given: argument set $\mathbf{A}$, argument A

5. Π_{argdoc} Is the document argumentative?

   Given: document d

6. Π_{argquery} Is the query argumentative?

   Given: query q

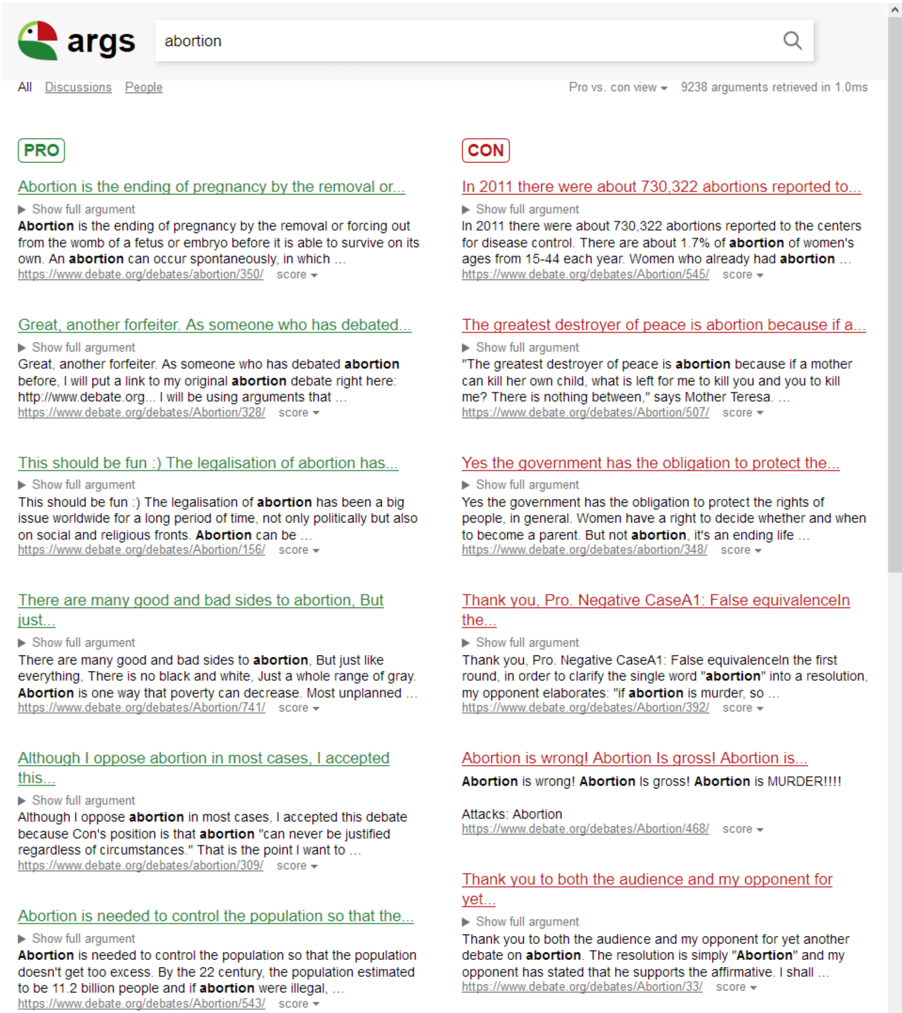
7. Π_{argsum} Summarize an argument.

   Given: argument A

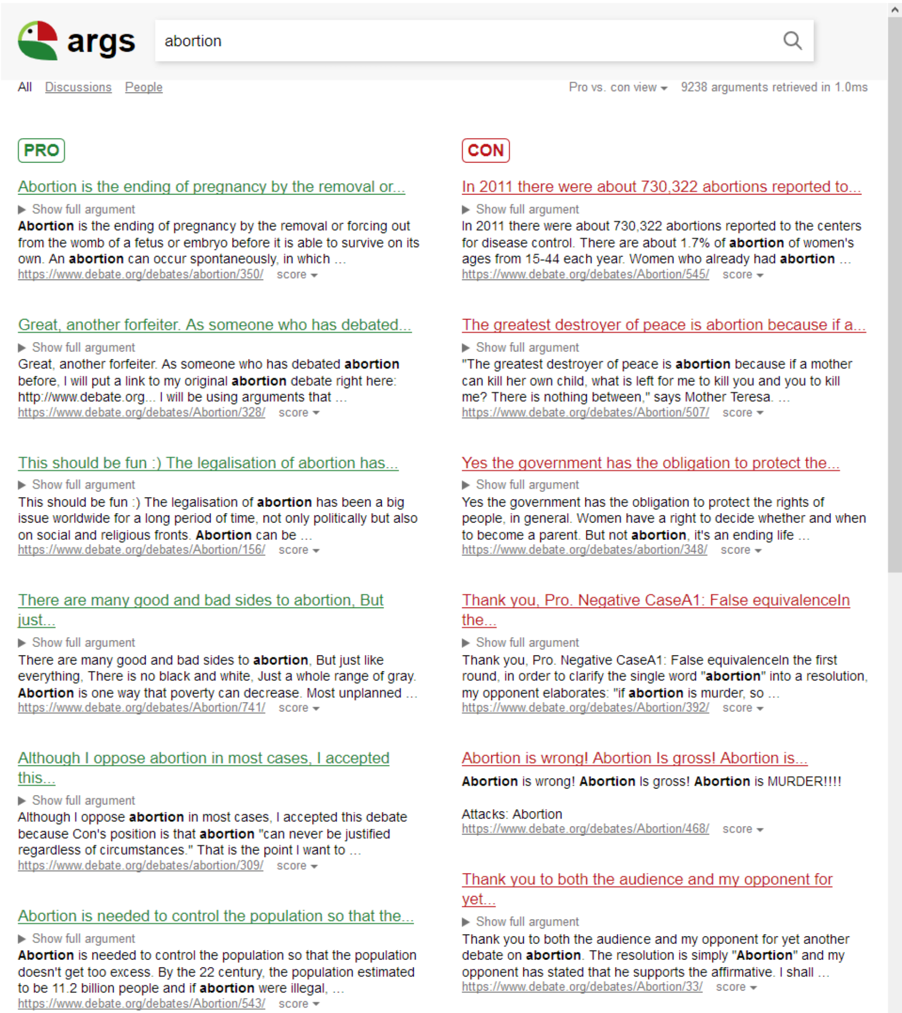
Notes:

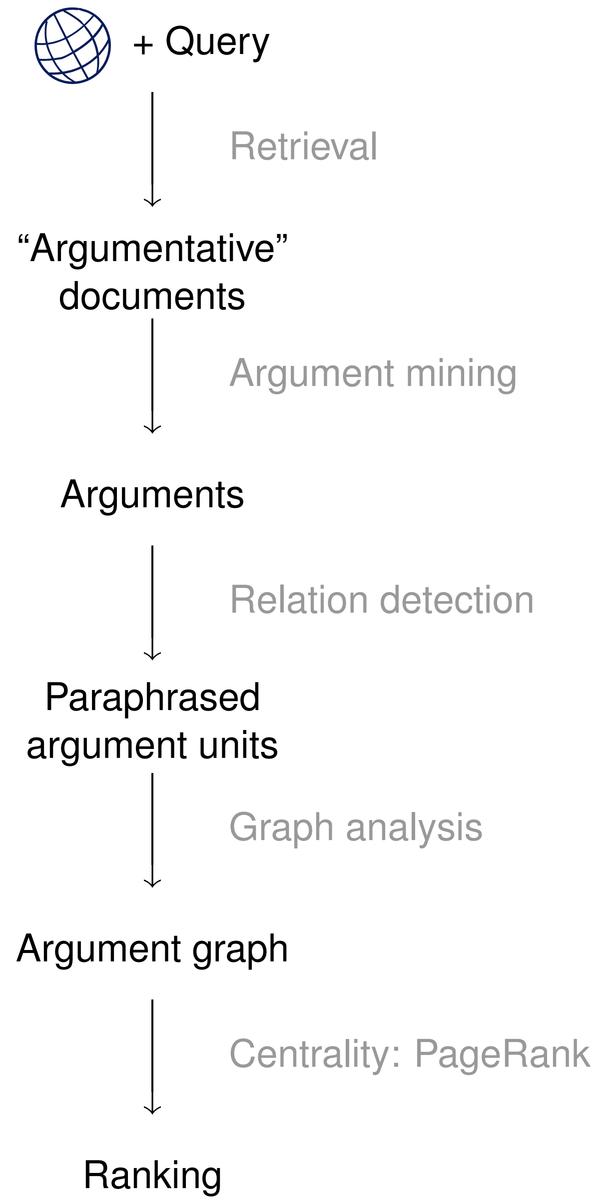
- Π_{counter} can be cast as Π_{rank} if the query is negated.
- Π_{argdoc} and Π_{argquery} are decision problems.
- Π_{counter} and Π_{sameside} can be cast as decision problems as well.
- Challenge: development of domain-independent or "topic-agnostic" approaches.

Argument Ranking I



Argument Ranking I

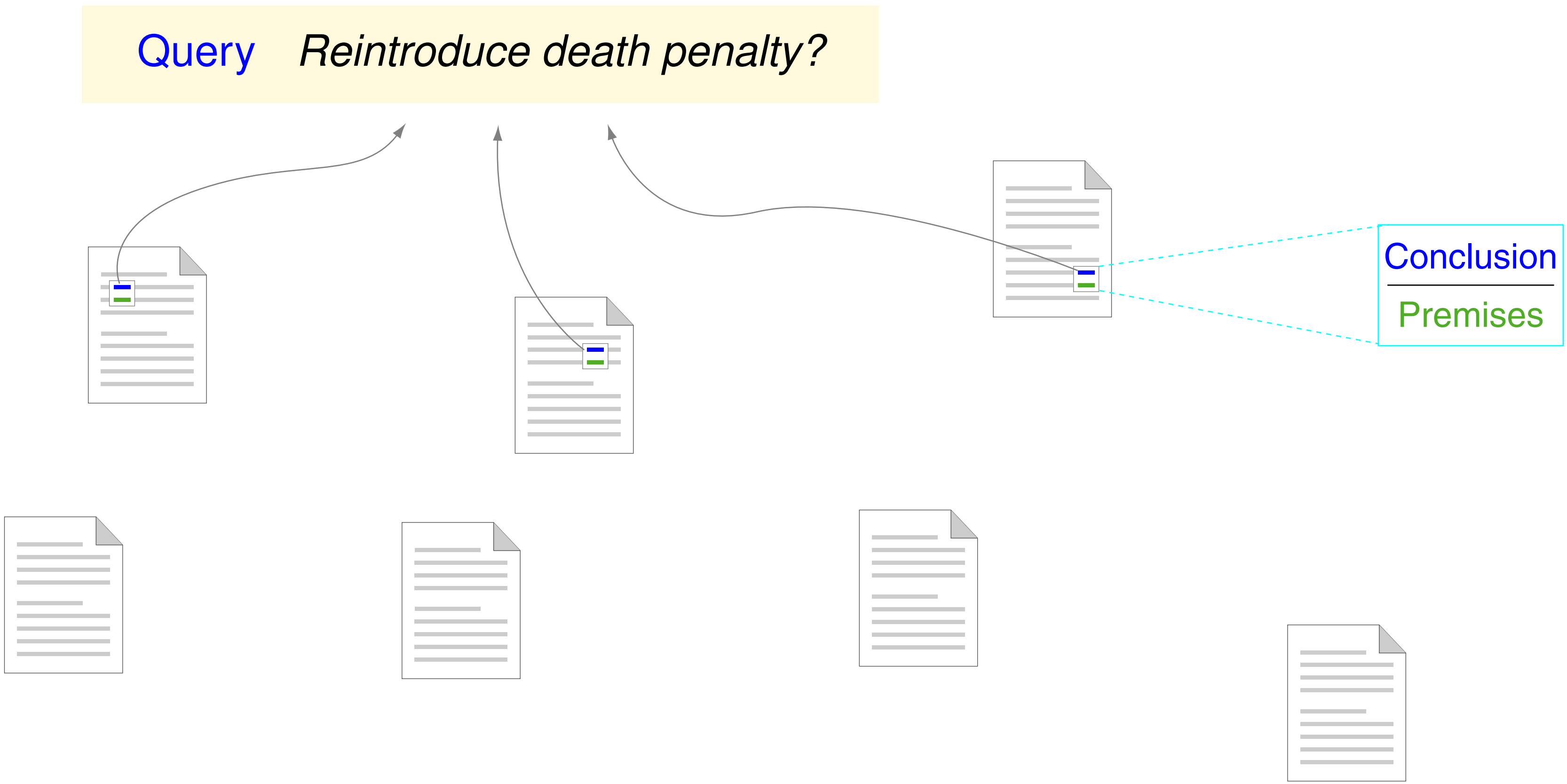




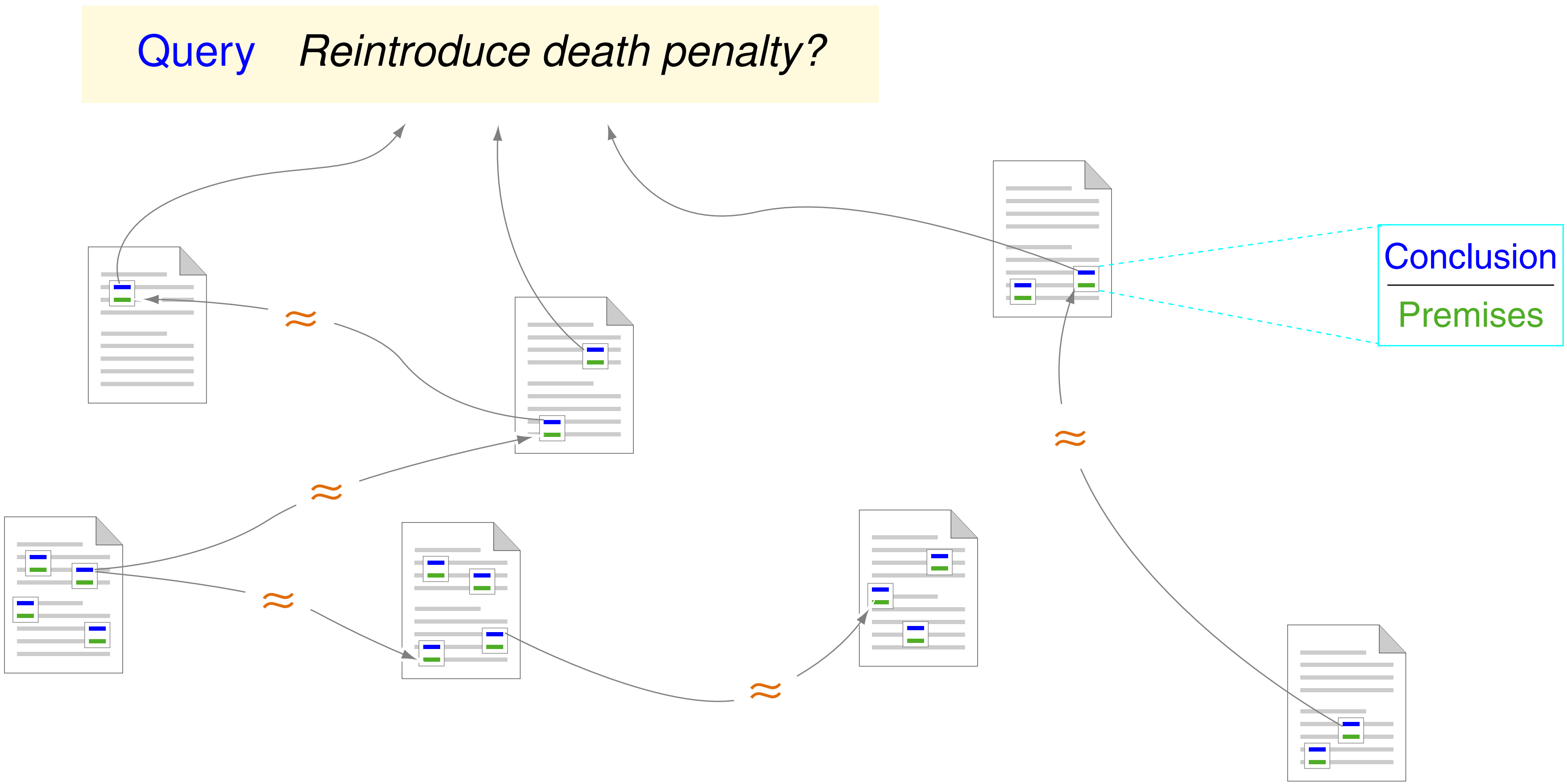
Argument Ranking I

Query *Reintroduce death penalty?*

Argument Ranking I

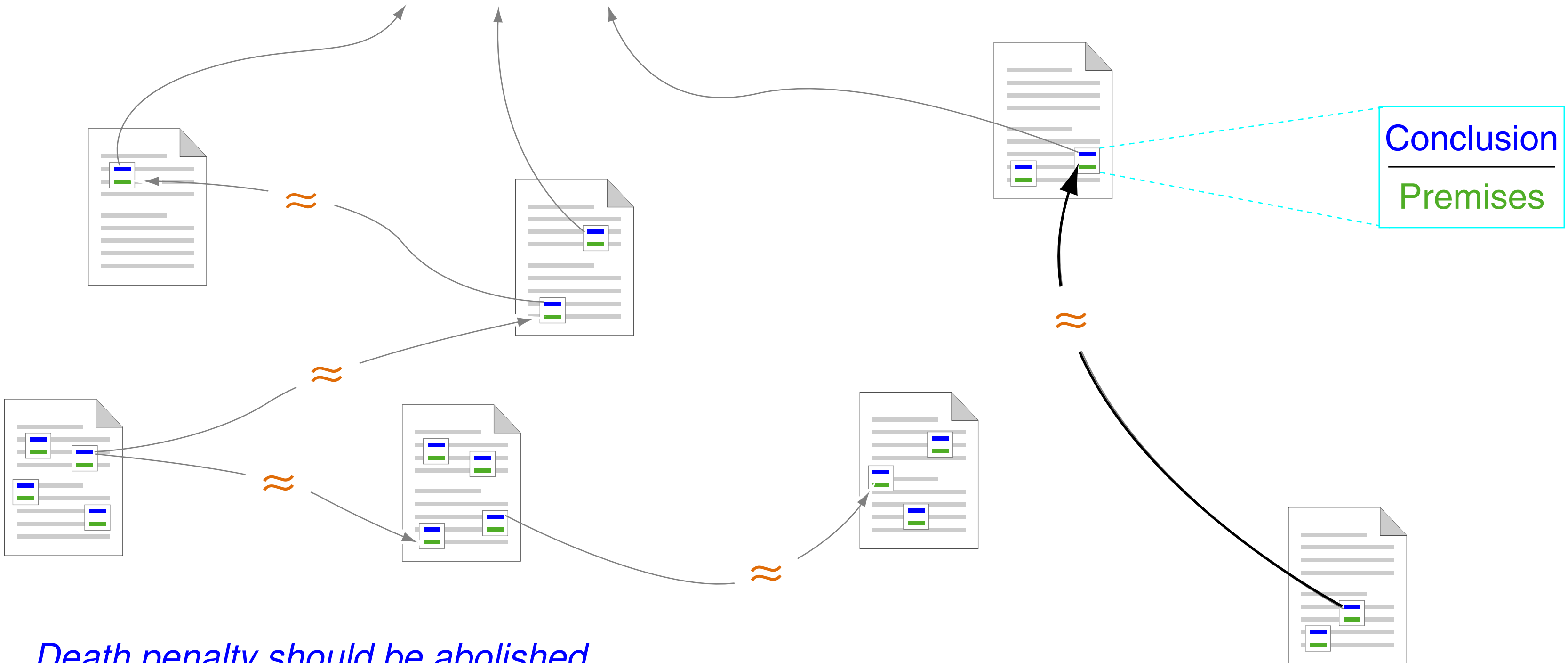


Argument Ranking I



Argument Ranking I

Query *Reintroduce death penalty?*



Death penalty should be abolished.

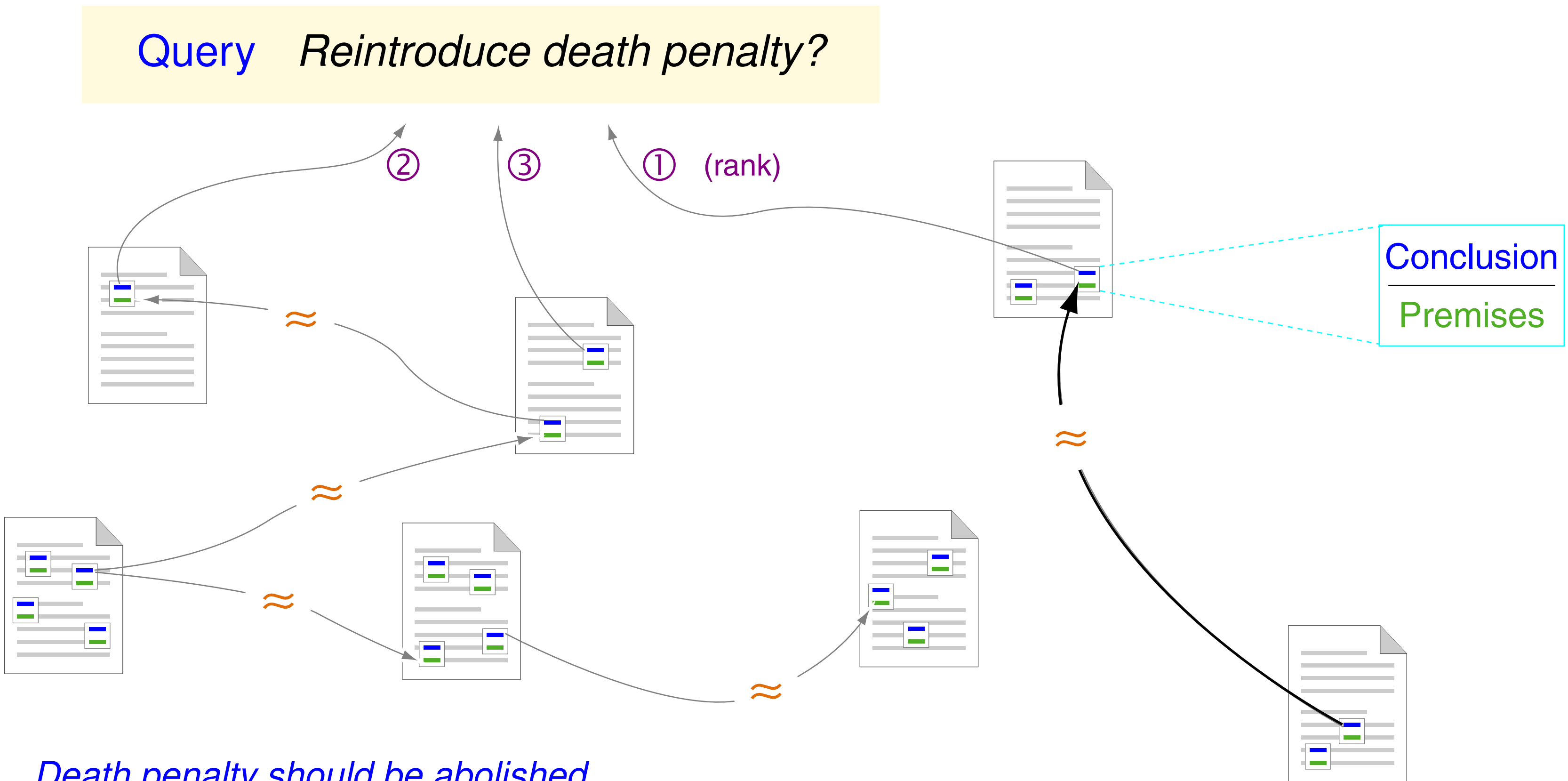
It does not prevent people from committing crimes.

≈ ←

The death penalty doesn't deter people from committing serious violent crimes.

A survey of the UN on the relation between the death penalty and homicide rates gave no support to the deterrent hypothesis.

Argument Ranking I



Death penalty should be abolished.

It does not prevent people from committing crimes.

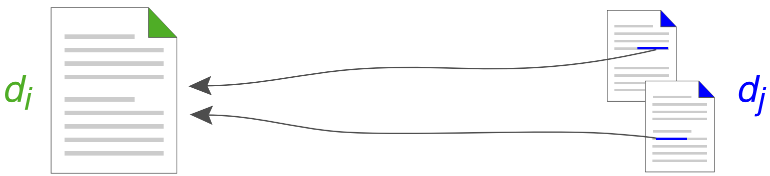
← ≈

The death penalty doesn't deter people from committing serious violent crimes.

A survey of the UN on the relation between the death penalty and homicide rates gave no support to the deterrent hypothesis.

Argument Ranking I

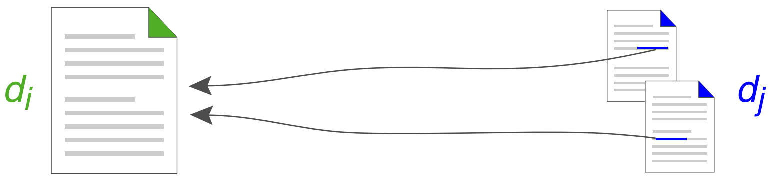
$$p(d_i) = (1-\alpha) \cdot \frac{1}{|D|} + \alpha \cdot \sum_j \frac{p(d_j)}{|D_j|}$$



Original PageRank [Page et al. 1999]

Argument Ranking I

$$p(d_i) = (1-\alpha)\cdot \boxed{\frac{1}{|D|}} + \alpha \cdot \boxed{\sum_j \frac{p(d_j)}{|D_j|}}$$

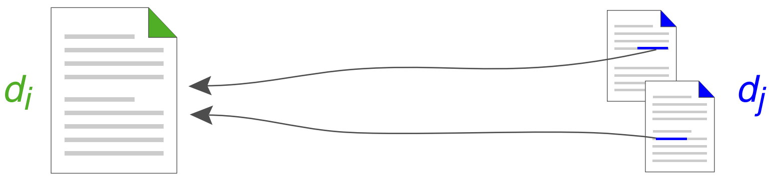


Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance

Argument Ranking I

$$p(d_i) = (1 - \alpha) \cdot \frac{1}{|D|} + \alpha \cdot \boxed{\sum_j \frac{p(d_j)}{|D_j|}}$$

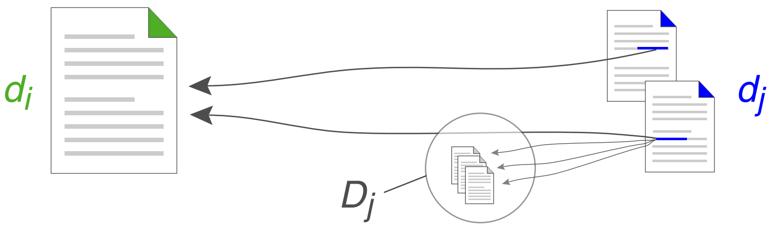


Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance
2. d_j links to d_i $\rightsquigarrow$ increase PageRank(d_i)

Argument Ranking I

$$p(d_i) = (1-\alpha)\cdot \frac{1}{|D|} + \alpha \cdot \boxed{\sum_j \frac{p(d_j)}{|D_j|}}$$

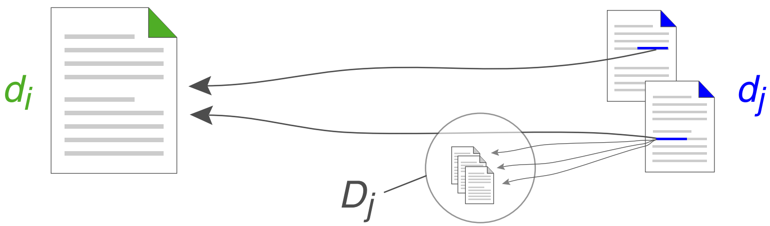


Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance
2. d_j links to d_i $\rightsquigarrow$ increase PageRank(d_i)
3. reward exclusive links

Argument Ranking I

$$p(d_i) = (1 - \alpha) \cdot \boxed{\frac{1}{|D|}} + \alpha \cdot \sum_j \frac{p(d_j)}{|D_j|}$$

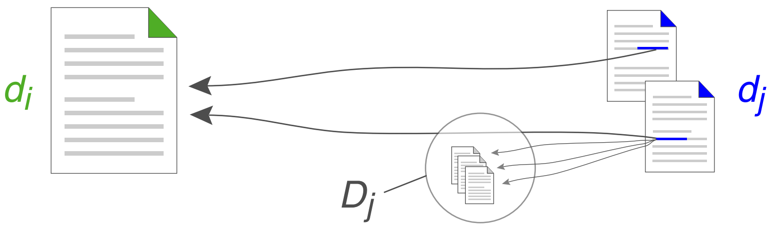


Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance
2. d_j links to d_i $\rightsquigarrow$ increase PageRank(d_i)
3. reward exclusive links
4. uniform ground relevances (sum to 1)

Argument Ranking I

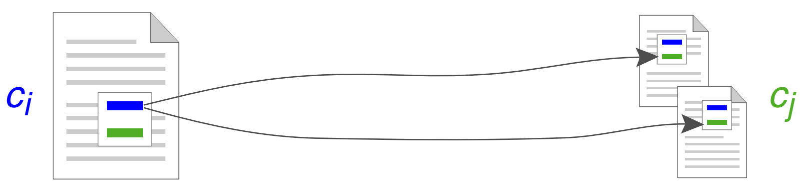
$$p(d_i) = (1-\alpha)\cdot\frac{1}{|D|} + \alpha\cdot\sum_j \frac{p(d_j)}{|D_j|}$$



Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance
2. d_j links to d_i $\rightsquigarrow$ increase PageRank(d_i)
3. reward exclusive links
4. uniform ground relevances (sum to 1)

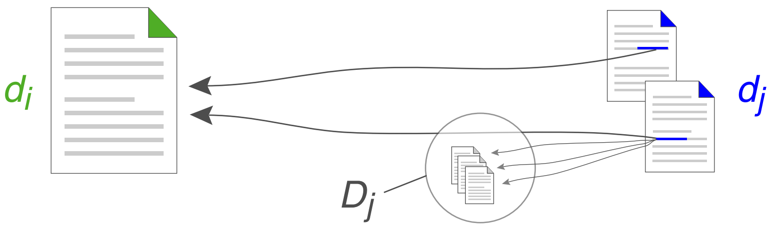
$$\hat{p}(c_i) = (1-\alpha)\cdot\frac{p(d_i)\cdot|D|}{|A|} + \alpha\cdot\sum_j \frac{\hat{p}(c_j)}{|P_j|}$$



ArgRank [Wachsmuth/Stein 2017]

Argument Ranking I

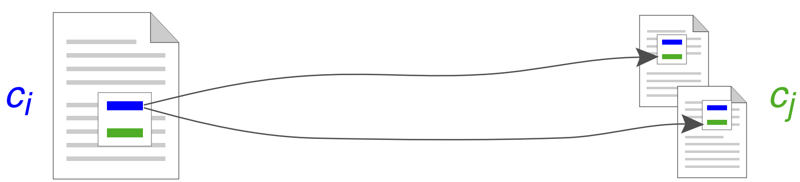
$$p(d_i) = (1-\alpha)\cdot\frac{1}{|D|} + \alpha\cdot\sum_j \frac{p(d_j)}{|D_j|}$$



Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance
2. d_j links to d_i $\leadsto$ increase PageRank(d_i)
3. reward exclusive links
4. uniform ground relevances (sum to 1)

$$\hat{p}(c_i) = (1-\alpha)\cdot\boxed{\frac{p(d_i)\cdot|D|}{|A|}} + \alpha\cdot\boxed{\sum_j \frac{\hat{p}(c_j)}{|P_j|}}$$

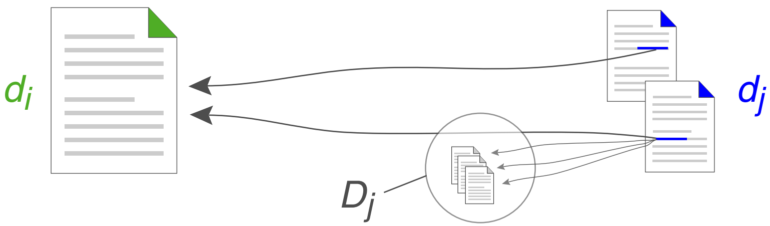


ArgRank [Wachsmuth/Stein 2017]

1. ground strength + recursive relevance

Argument Ranking I

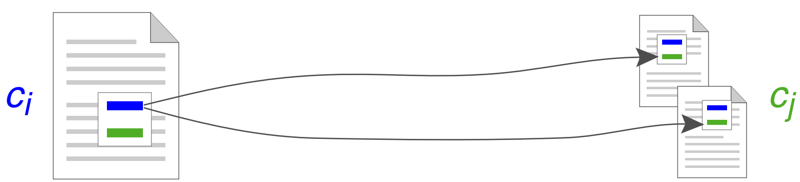
$$p(d_i) = (1-\alpha)\cdot \frac{1}{|D|} + \alpha \cdot \sum_j \frac{p(d_j)}{|D_j|}$$



Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance
2. d_j links to d_i $\leadsto$ increase PageRank(d_i)
3. reward exclusive links
4. uniform ground relevances (sum to 1)

$$\hat{p}(c_i) = (1-\alpha)\cdot \frac{p(d_i)\cdot|D|}{|A|} + \alpha\cdot \boxed{\sum_j \frac{\hat{p}(c_j)}{|P_j|}}$$

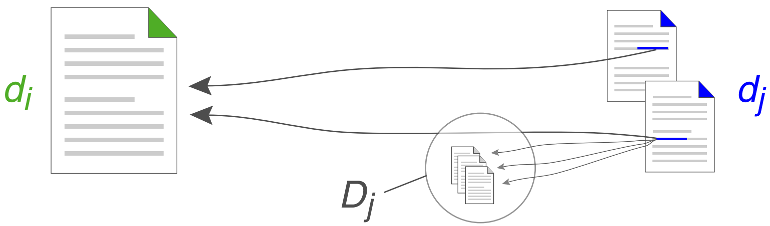


ArgRank [Wachsmuth/Stein 2017]

1. ground strength + recursive relevance
2. c_i premise for c_j $\leadsto$ increase ArgRank(c_i)

Argument Ranking I

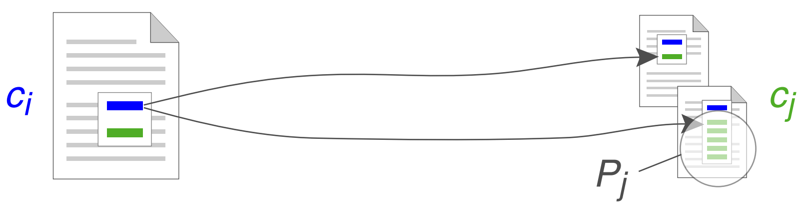
$$p(d_i) = (1-\alpha)\cdot \frac{1}{|D|} + \alpha \cdot \sum_j \frac{p(d_j)}{|D_j|}$$



Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance
2. d_j links to d_i $\rightsquigarrow$ increase PageRank(d_i)
3. reward exclusive links
4. uniform ground relevances (sum to 1)

$$\hat{p}(c_i) = (1-\alpha)\cdot \frac{p(d_i)\cdot|D|}{|A|} + \alpha \cdot \boxed{\sum_j \frac{\hat{p}(c_j)}{|P_j|}}$$

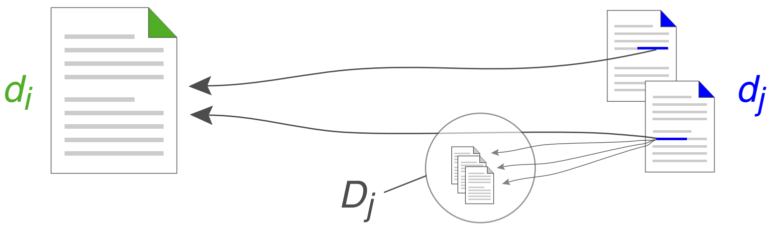


ArgRank [Wachsmuth/Stein 2017]

1. ground strength + recursive relevance
2. c_i premise for c_j $\rightsquigarrow$ increase ArgRank(c_i)
3. reward exclusive premises

Argument Ranking I

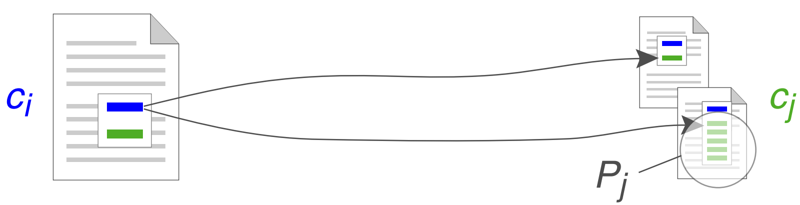
$$p(d_i) = (1-\alpha) \cdot \frac{1}{|D|} + \alpha \cdot \sum_j \frac{p(d_j)}{|D_j|}$$



Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance
2. d_j links to d_i $\rightsquigarrow$ increase PageRank(d_i)
3. reward exclusive links
4. uniform ground relevances (sum to 1)

$$\hat{p}(c_i) = (1-\alpha) \cdot \boxed{\frac{p(d_i) \cdot |D|}{|A|}} + \alpha \cdot \sum_j \frac{\hat{p}(c_j)}{|P_j|}$$

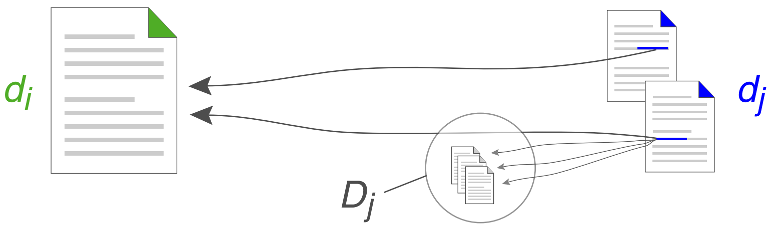


ArgRank [Wachsmuth/Stein 2017]

1. ground strength + recursive relevance
2. c_i premise for c_j $\rightsquigarrow$ increase ArgRank(c_i)
3. reward exclusive premises
4. ground strength $\sim$ PageRank

Argument Ranking I

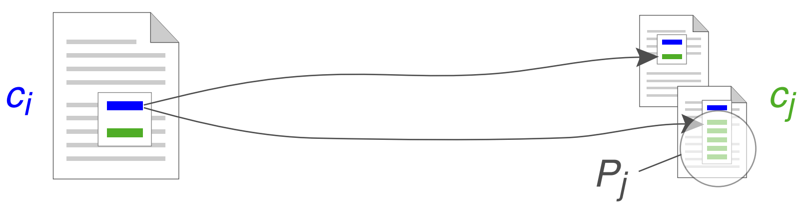
$$p(d_i) = (1-\alpha)\cdot\frac{1}{|D|} + \alpha\cdot\sum_j \frac{p(d_j)}{|D_j|}$$



Original PageRank [Page et al. 1999]

1. ground relevance + recursive relevance
2. d_j links to d_i $\rightsquigarrow$ increase PageRank(d_i)
3. reward exclusive links
4. uniform ground relevances (sum to 1)

$$\hat{p}(c_i) = (1-\alpha)\cdot\frac{p(d_i)\cdot|D|}{|A|} + \alpha\cdot\sum_j \frac{\hat{p}(c_j)}{|P_j|}$$



ArgRank [Wachsmuth/Stein 2017]

1. ground strength + recursive relevance
2. c_i premise for c_j $\rightsquigarrow$ increase ArgRank(c_i)
3. reward exclusive premises
4. ground strength $\sim$ PageRank

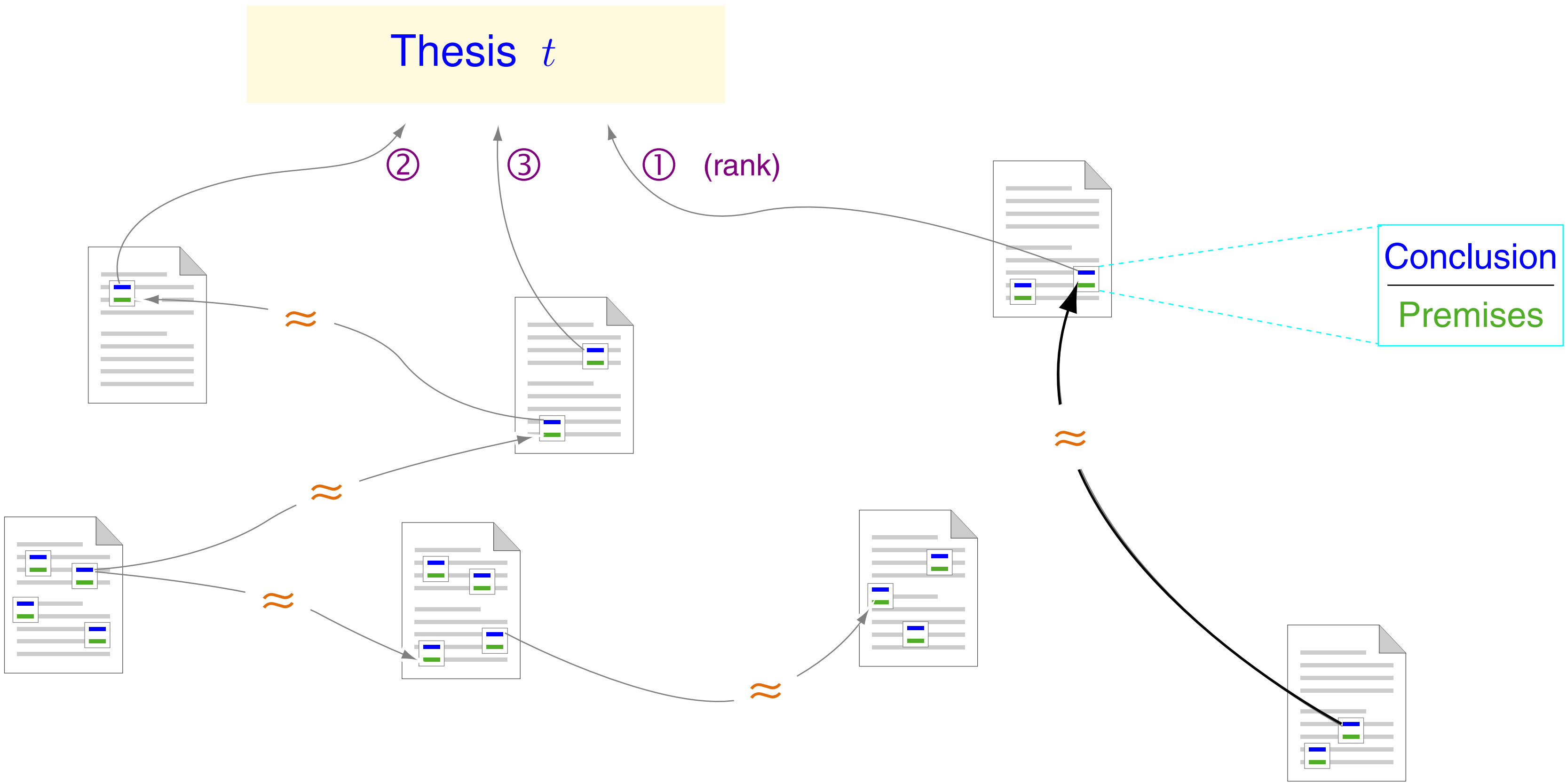
"Reversal of Evidence"

PageRank: Author cannot enforce links to her web page.

ArgRank: Author cannot enforce use of her argument.

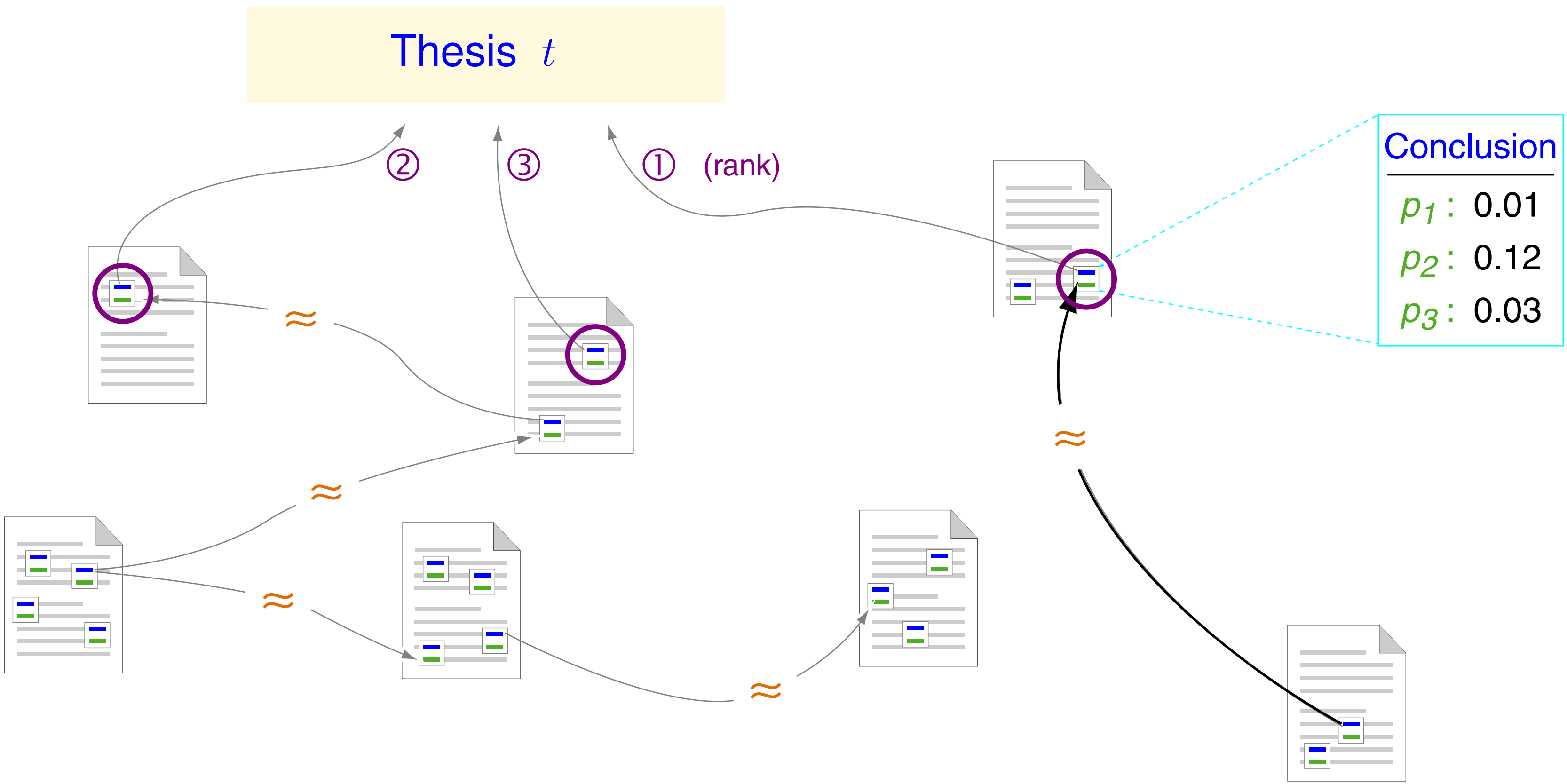
Argument Ranking I

From Premise Scores to Argument Ranks



Argument Ranking I

From Premise Scores to Argument Ranks



How to weigh the premise scores of the matching arguments?

(maximum, average, etc.)

Argument Ranking I

Case Study: Graph Construction

Construction of a raw graph using 57 corpora from the Argument Web :

28 875 Argument units, used in . . .

17 877 Arguments

Processing steps towards an argument graph:

3 113 Conclusions with $\geq$ 1 argument, where . . .

498 have multiple premises, from which . . .

70 have a relevant claim, from which . . .

32 are used in 110 intelligible arguments.

Argument Ranking I

Case Study: Graph Construction

Construction of a raw graph using 57 corpora from the Argument Web :

28 875 Argument units, used in . . .

17 877 Arguments

Processing steps towards an argument graph:

3 113 Conclusions with $\geq$ 1 argument, where . . .

498 have multiple premises, from which . . .

70 have a relevant claim, from which . . .

32 are used in 110 intelligible arguments.

Acquisition of a ranking ground truth:

- 7 experts from NLP and IR ranked all arguments (110) for each conclusion (32)
- $\tau = 0.59$ as highest agreement between two experts (mean: $\tau = 0.36$)

Argument Ranking I

Case Study: Results

Ranking approach	Premise score computation				Best
	Minimum	Average	Maximum	Sum	
	τ	τ	τ	τ	τ
1. ArgRank	**0.01**	0.02	0.11	**0.28**	**0.28**
2. Frequency	–0.10	–0.03	–0.01	0.10	0.10
3. Similarity	–0.13	–0.05	0.01	0.02	0.02
4. Sentiment	**0.01**	**0.11**	**0.12**	0.12	0.12
5. Most premises	-	-	-	-	0.19
6. Random	-	-	-	-	0.00

Approach 1: An argument's relevance corresponds to the ArgRank of its premises.

Argument Ranking I

Case Study: Results

Ranking approach	**Premise score computation**				**Best**
	Minimum	Average	Maximum	Sum	
	τ	τ	τ	τ	τ
1. ArgRank	**0.01**	0.02	0.11	**0.28**	**0.28**
2. Frequency	–0.10	–0.03	–0.01	0.10	0.10
3. Similarity	–0.13	–0.05	0.01	0.02	0.02
4. Sentiment	**0.01**	**0.11**	**0.12**	0.12	0.12
5. Most premises	-	-	-	-	0.19
6. Random	-	-	-	-	0.00

Approach 2: An argument's relevance corresponds to the frequency of its premises in the graph.

Argument Ranking I

Case Study: Results

Ranking approach	Premise score computation				Best
	Minimum	Average	Maximum	Sum	
	τ	τ	τ	τ	τ
1. ArgRank	**0.01**	0.02	0.11	**0.28**	**0.28**
2. Frequency	–0.10	–0.03	–0.01	0.10	0.10
3. Similarity	–0.13	–0.05	0.01	0.02	0.02
4. Sentiment	**0.01**	**0.11**	**0.12**	0.12	0.12
5. Most premises	-	-	-	-	0.19
6. Random	-	-	-	-	0.00

Approach 3: An argument's relevance corresponds to the Jaccard similarity of its premises to its conclusion.

Argument Ranking I

Case Study: Results

Ranking approach	Premise score computation				Best
	Minimum	Average	Maximum	Sum	
	τ	τ	τ	τ	τ
1. ArgRank	**0.01**	0.02	0.11	**0.28**	**0.28**
2. Frequency	–0.10	–0.03	–0.01	0.10	0.10
3. Similarity	–0.13	–0.05	0.01	0.02	0.02
4. Sentiment	**0.01**	**0.11**	**0.12**	0.12	0.12
5. Most premises	-	-	-	-	0.19
6. Random	-	-	-	-	0.00

Approach 4: An argument's relevance corresponds to the positivity of its words in the premises according to SentiWordNet.

Argument Ranking I

Case Study: Results

Ranking approach	Premise score computation				Best
	Minimum	Average	Maximum	Sum	
	τ	τ	τ	τ	τ
1. ArgRank	**0.01**	0.02	0.11	**0.28**	**0.28**
2. Frequency	–0.10	–0.03	–0.01	0.10	0.10
3. Similarity	–0.13	–0.05	0.01	0.02	0.02
4. Sentiment	**0.01**	**0.11**	**0.12**	0.12	0.12
5. Most premises	-	-	-	-	0.19
6. Random	-	-	-	-	0.00

Approach 5: An argument's relevance corresponds to its number of premises.

Argument Ranking I

Case Study: Results

Ranking approach	Premise score computation				Best
	Minimum	Average	Maximum	Sum	
	τ	τ	τ	τ	τ
1. ArgRank	**0.01**	0.02	0.11	**0.28**	**0.28**
2. Frequency	–0.10	–0.03	–0.01	0.10	0.10
3. Similarity	–0.13	–0.05	0.01	0.02	0.02
4. Sentiment	**0.01**	**0.11**	**0.12**	0.12	0.12
5. Most premises	-	-	-	-	0.19
6. Random	-	-	-	-	0.00

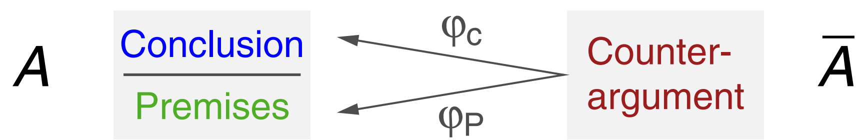
Approach 6: The relevance is decided randomly.

Argument Ranking II

Argument Ranking II

Idea: Given an argument A, the best counterargument $\overline{A}^*$ employs premises that are similar wrt. topic, but takes the opposite stance.

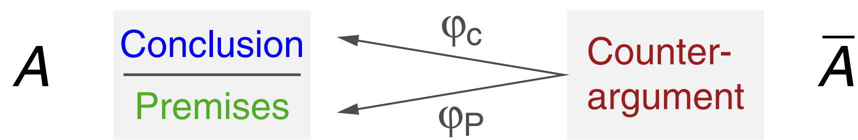
➜ Consider both similarities to the premises and conclusion [Walton 2009] :



Argument Ranking II

Idea: Given an argument A, the best counterargument $\overline{A}^*$ employs premises that are similar wrt. topic, but takes the opposite stance.

➜ Consider both similarities to the premises and conclusion [Walton 2009] :



How to compute these similarities?

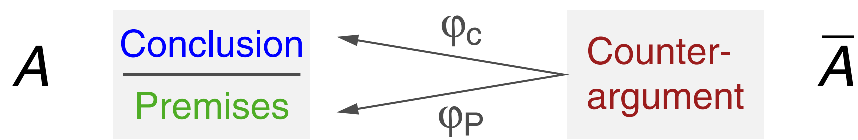
How to combine these similarities?

(= What is a sensible hypothesis space of promising model functions?)

Argument Ranking II

Idea: Given an argument A, the best counterargument $\overline{A}^*$ employs premises that are similar wrt. topic, but takes the opposite stance.

➜ Consider both similarities to the premises and conclusion [Walton 2009] :



Proposed model function to rank counterarguments [Wachsmuth et al., 2018] :

$$R(A,\overline{A}) = \alpha \cdot \underbrace{\left(\varphi\text{conclusion} \circ \varphi\text{Premises}\right)}_{\text{topic similarity} \,\rightarrow\, \max} - (1-\alpha) \cdot \underbrace{\left(\varphi\text{conclusion} \circ \varphi\text{Premises}\right)}_{\text{stance similarity} \,\rightarrow\, \min}$$

where

φ combines both word and embedding similarities

$\circ \in \{\min, \max, +, *\}$

$\alpha \in [0;1]$

Argument Ranking II

Corpus and Analysis

Theme	Debates	Points	Counters
Culture	46	278	278
Digital freedoms	48	341	341
Economy	95	590	588
⋮			
Sport	23	130	130
$\sum$	**1069**	**6779**	**6753**

Corpus:

- based on the iDebate.org portal
- Download: ArguAna Counterargs

Argument Ranking II

Corpus and Analysis

Theme	Debates	Points	Counters
Culture	46	278	278
Digital freedoms	48	341	341
Economy	95	590	588
⋮			
Sport	23	130	130
$\sum$	**1069**	**6779**	**6753**

Corpus:

- based on the iDebate.org portal
- Download: ArguAna Counterargs

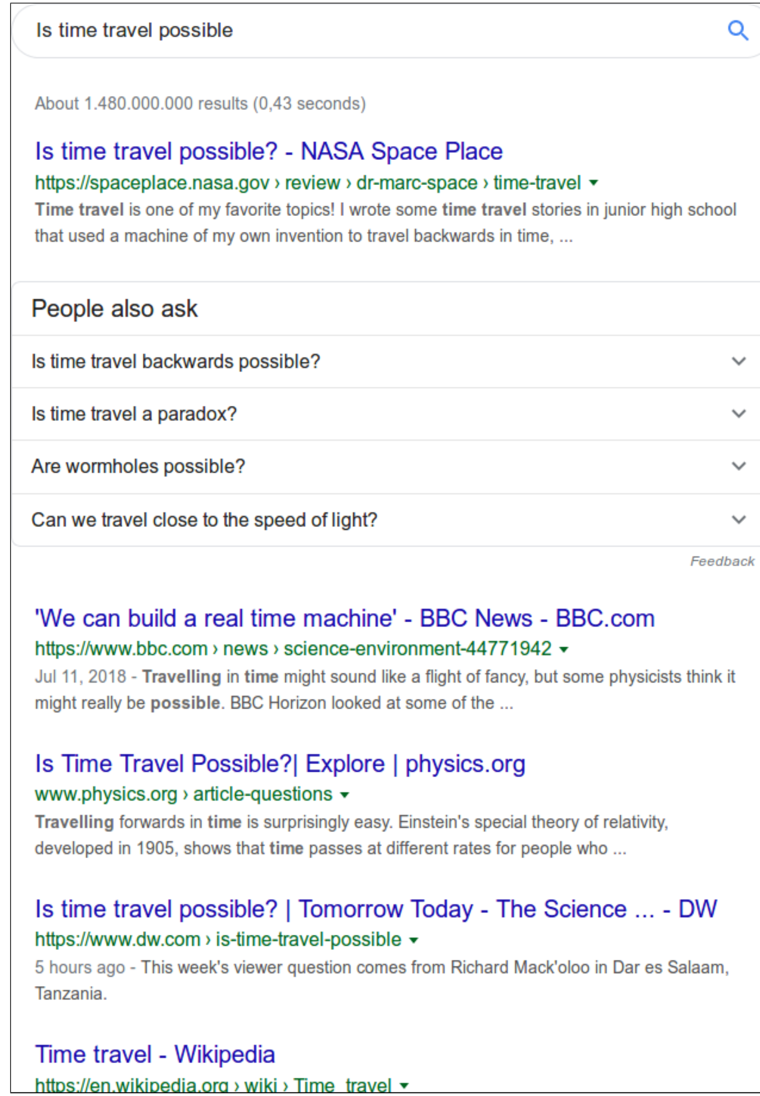
Retrieval experiments (selected results) :

Find the best counterargument within . . .	True-to-false ratio	Accuracy*
all counters of the same debate	1 : 3	0.75
all counters of the same theme	1 : 136	0.54
all arguments of the entire portal	1 : 2 800	0.32

* The parameters for $R(A, \overline{A})$ were determined by a systematic ranking analysis.

Argument Search Engines

Vision of Argument Search



Argument Search Engines*

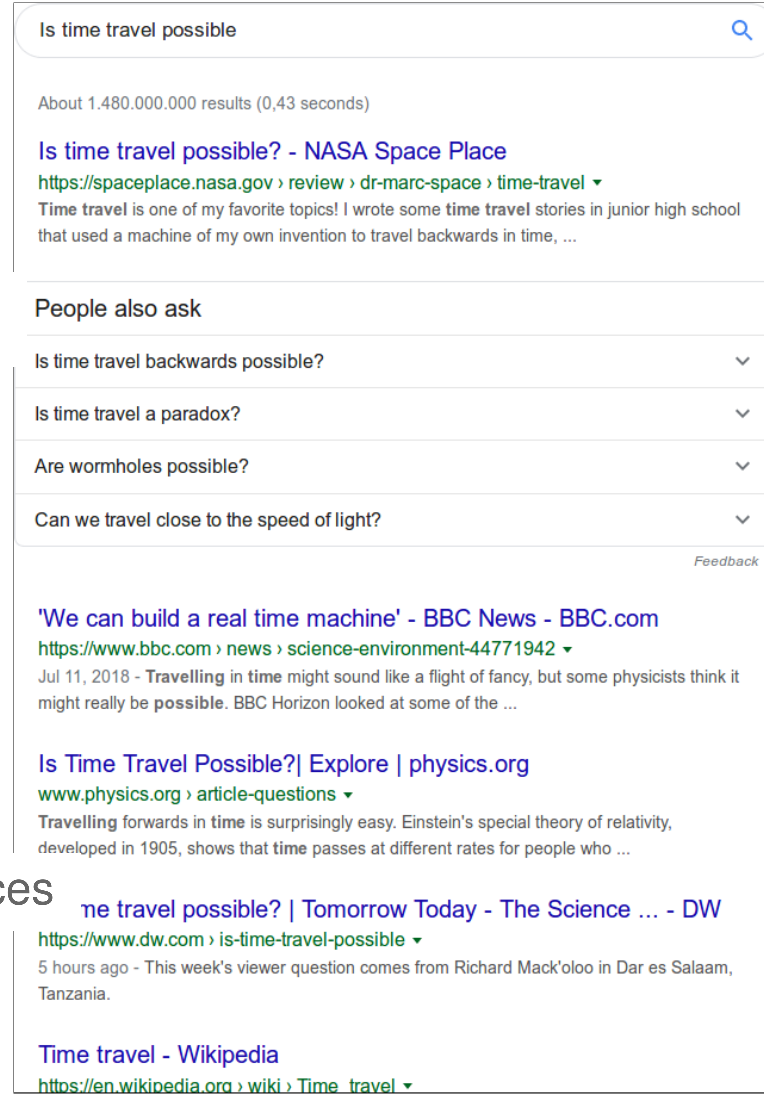
Vision of Argument Search

Arguments in future web search:

- support forming opinions
- make it easy to find relevant arguments
- deliberation: learn about other views
- education: learn to debate

Search results should . . .

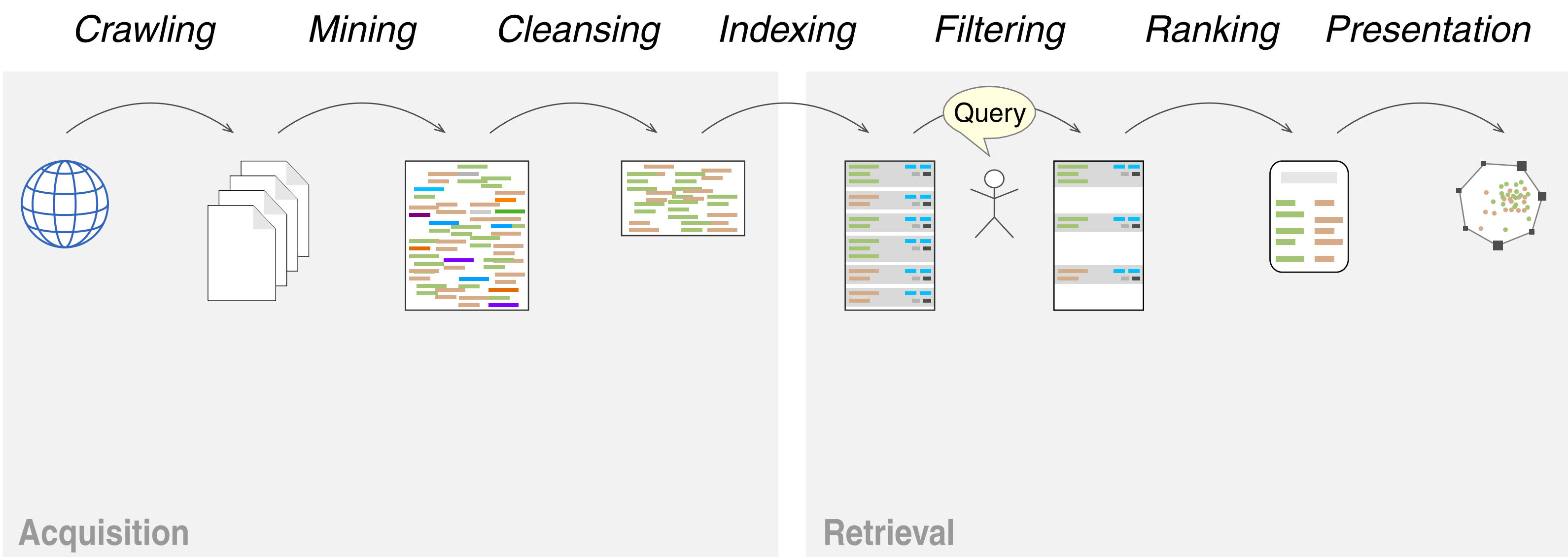
- rank the best arguments highest
- cover diverse aspects
- cover reliable and heterogeneous sources
- be up-to-the-minute
- be traceable and evaluable



* Wachsmuth: Argumentation Retrieval and Analysis. IR Autumn School ASIRF (2018).

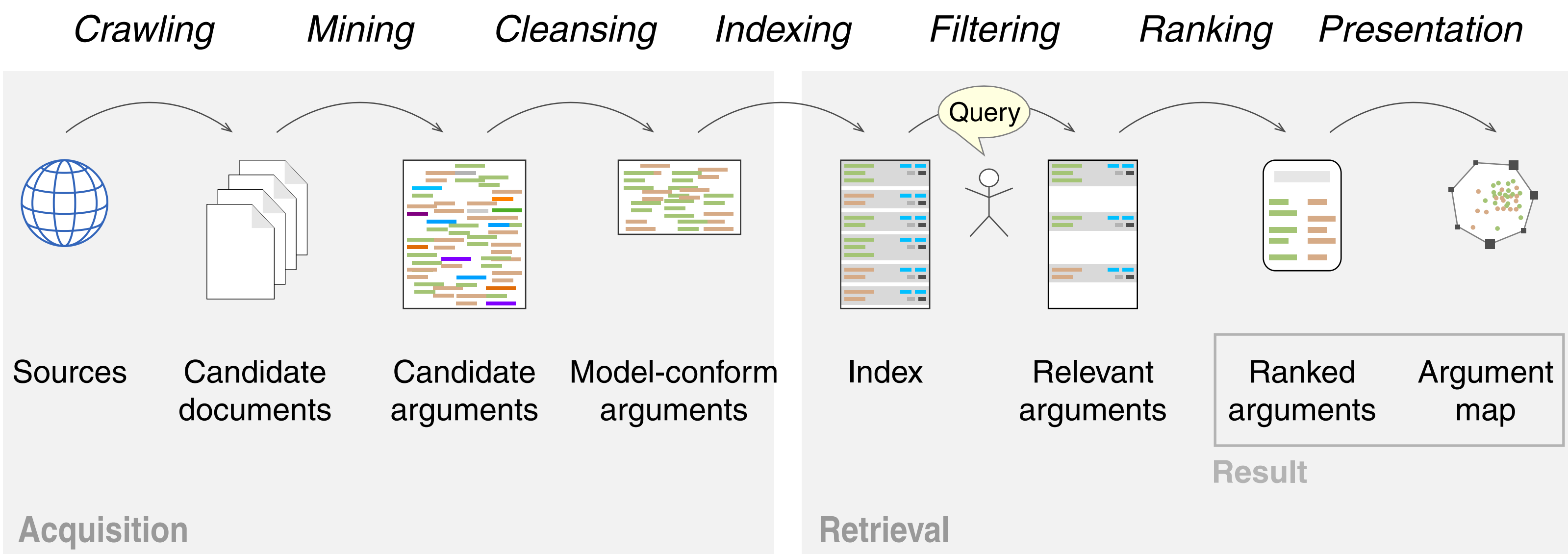
Argument Search Engines

Basic Elements and Process



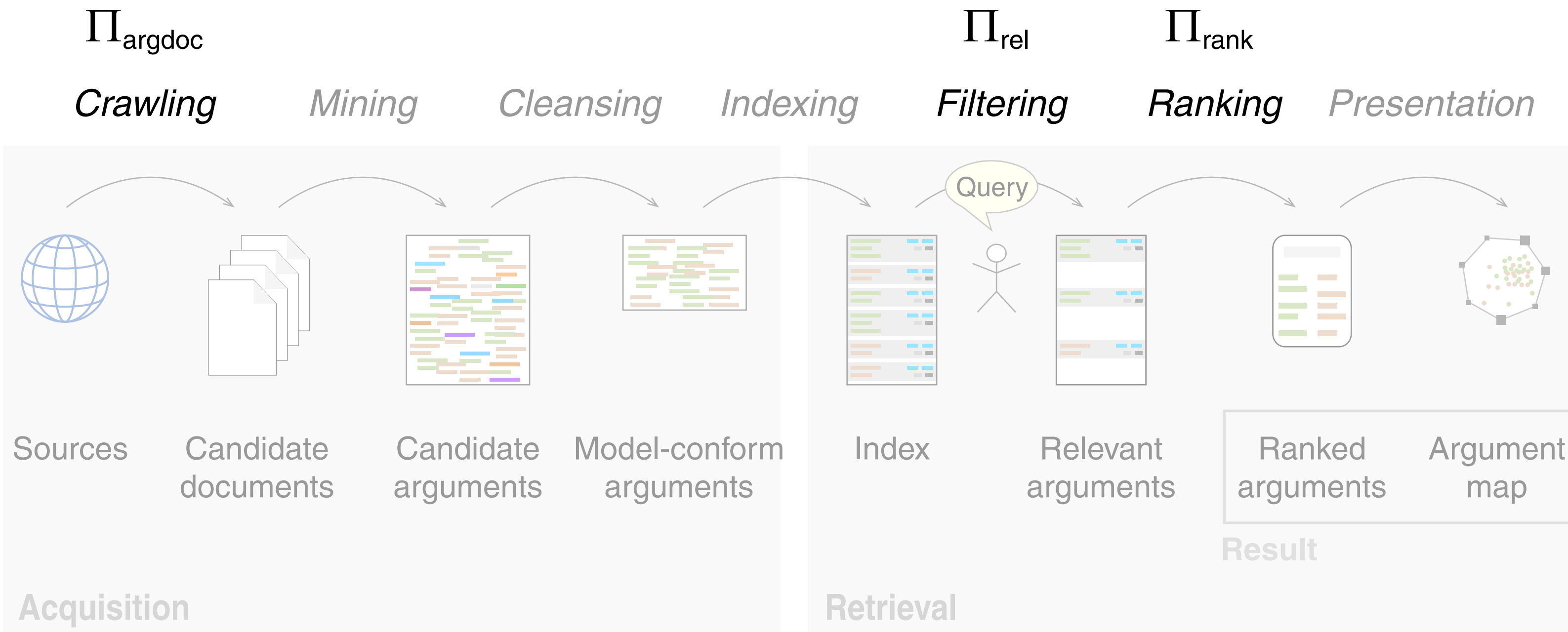
Argument Search Engines

Basic Elements and Process

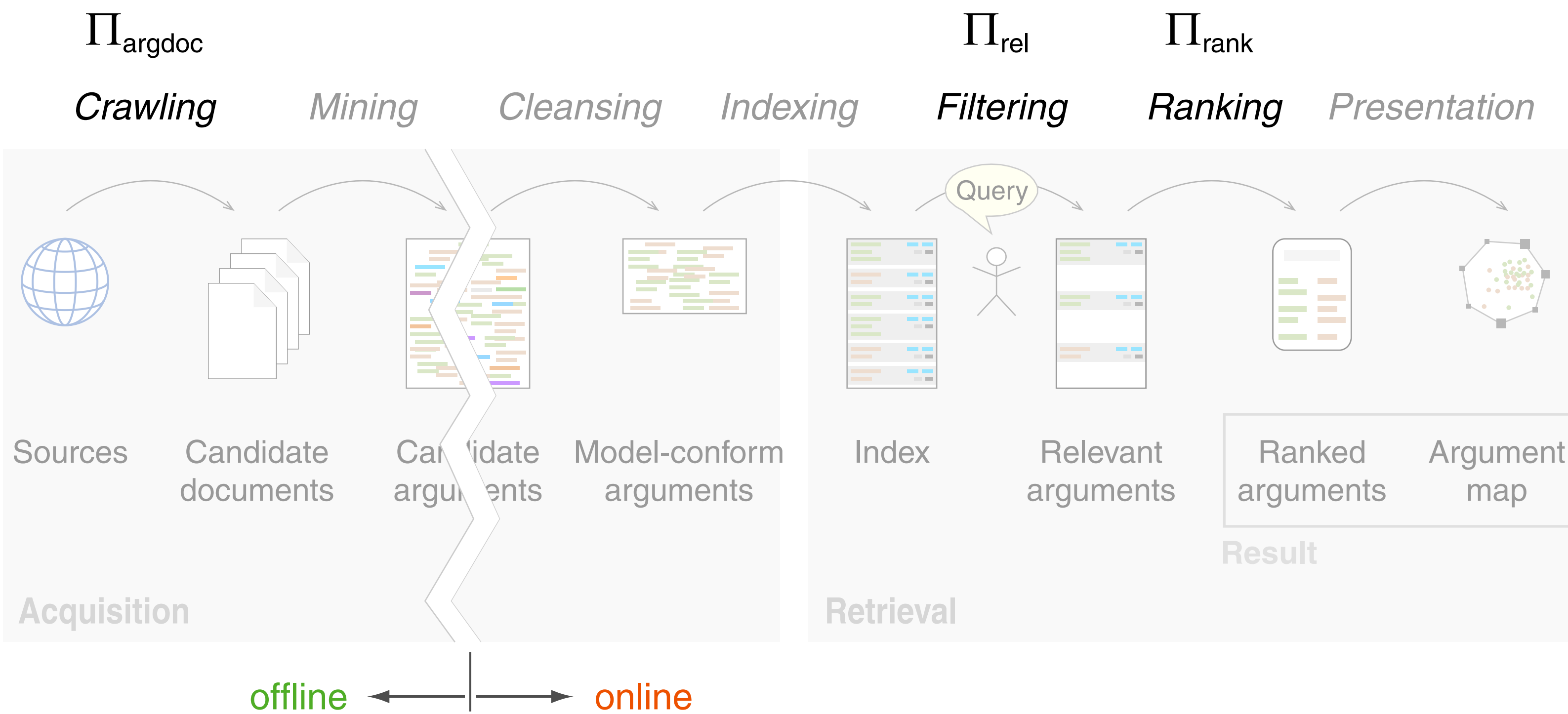


Argument Search Engines

Basic Elements and Process



Argument Search Engines
Basic Elements and Process



Acquisition paradigm [Ajjour et al. 2019] :

- distribution of processing steps regarding offline time and online time
- tradeoff between precision, recall, and topicality

Argument Search Engines

Acquisition Resources

Leverage effort*	Resource type	Examples
very low	Technology	
low	Corpora	
medium	Debate portals	
high	Discussion pages	
very high	Articles	

* Estimated effort / expertise to exploit a resource of the respective type within own research.

Argument Search Engines

Acquisition Resources

Leverage effort*	Resource type		Examples
very low	Technology	Visual inspection Acquisition, Tagging	Argument Web Truthmapping
low	Corpora		
medium	Debate portals		
high	Discussion pages		
very high	Articles		

* Estimated effort / expertise to exploit a resource of the respective type within own research.

Argument Search Engines

Acquisition Resources

Leverage effort*	Resource type		Examples
very low	Technology	Visual inspection Acquisition, Tagging	Argument Web Truthmapping
low	Corpora	Argumentative structure analysis Argumentation quality analysis Stance detection	AIFdb data IBM Debater data UKP data Webis data
medium	Debate portals		
high	Discussion pages		
very high	Articles		

* Estimated effort / expertise to exploit a resource of the respective type within own research.

Argument Search Engines

Acquisition Resources

Leverage effort*	Resource type		Examples
very low	Technology	Visual inspection Acquisition, Tagging	Argument Web Truthmapping
low	Corpora	Argumentative structure analysis Argumentation quality analysis Stance detection	AIFdb data IBM Debater data UKP data Webis data
medium	Debate portals	English	Kialo idebate Debatepedia
high	Discussion pages		
very high	Articles		

* Estimated effort / expertise to exploit a resource of the respective type within own research.

Argument Search Engines
Acquisition Resources

Leverage effort*	Resource type		Examples
very low	Technology	Visual inspection Acquisition, Tagging	Argument Web Truthmapping
low	Corpora	Argumentative structure analysis Argumentation quality analysis Stance detection	AIFdb data IBM Debater data UKP data Webis data
medium	Debate portals	English	Kialo idebate Debatepedia
high	Discussion pages	Focus on persuasion Controversial issues Focus on deliberation	changemyview reddit WikiTalk
very high	Articles		

* Estimated effort / expertise to exploit a resource of the respective type within own research.

Argument Search Engines

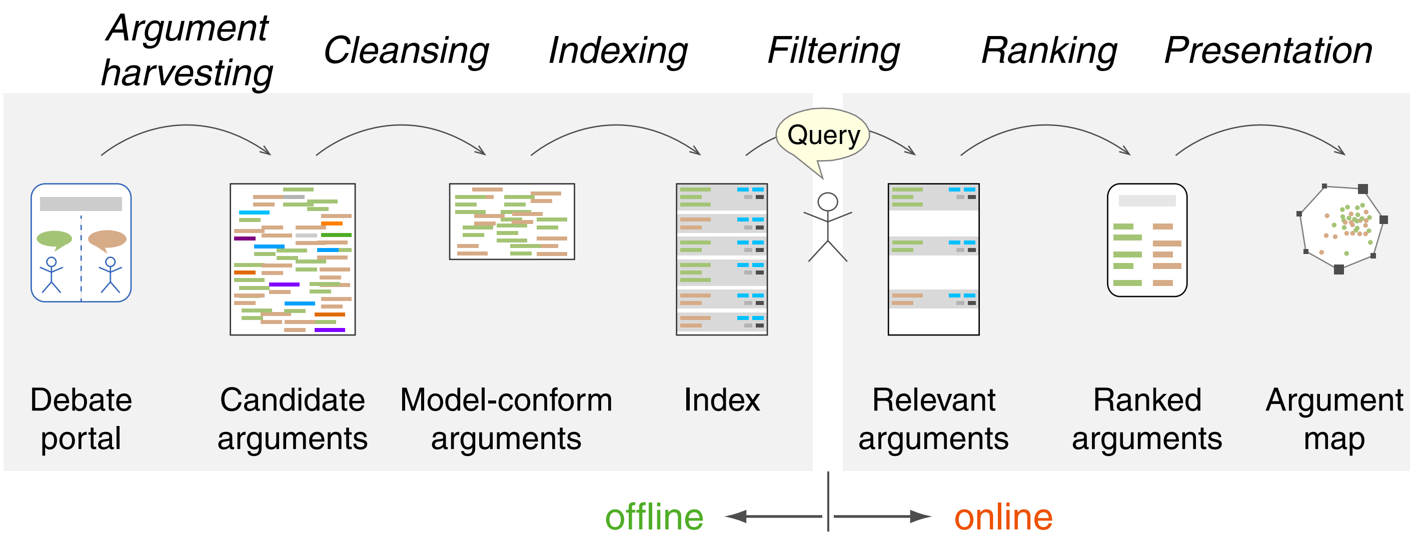
Acquisition Resources

Leverage effort*	Resource type		Examples
very low	Technology	Visual inspection Acquisition, Tagging	Argument Web Truthmapping
low	Corpora	Argumentative structure analysis Argumentation quality analysis Stance detection	AIFdb data IBM Debater data UKP data Webis data
medium	Debate portals	English	Kialo idebate Debatepedia
high	Discussion pages	Focus on persuasion Controversial issues Focus on deliberation	changemyview reddit WikiTalk
very high	Articles	Editorials, Essays Legal Scientific publications	New York Times ACL anthology

* Estimated effort / expertise to exploit a resource of the respective type within own research.

Argument Search Engines

Acquisition Paradigms: (a) args.me [Demo]



- Research focus: argument ranking
- Supervision level: medium (distantly supervised)

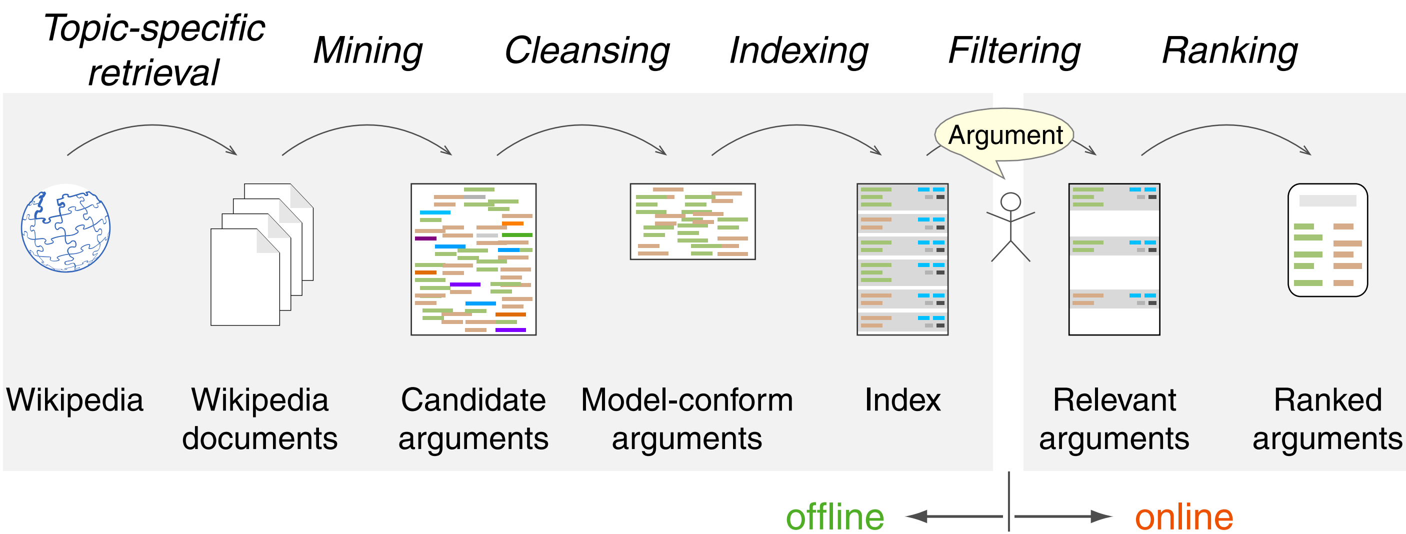
→ Effectiveness profile: high precision, low recall

→ Stance balance: guaranteed

→ Efficiency: high

Argument Search Engines

Acquisition Paradigms: (b) IBM Debater [Project]



- Research focus: debating technology
- Supervision level: medium (recognized source)

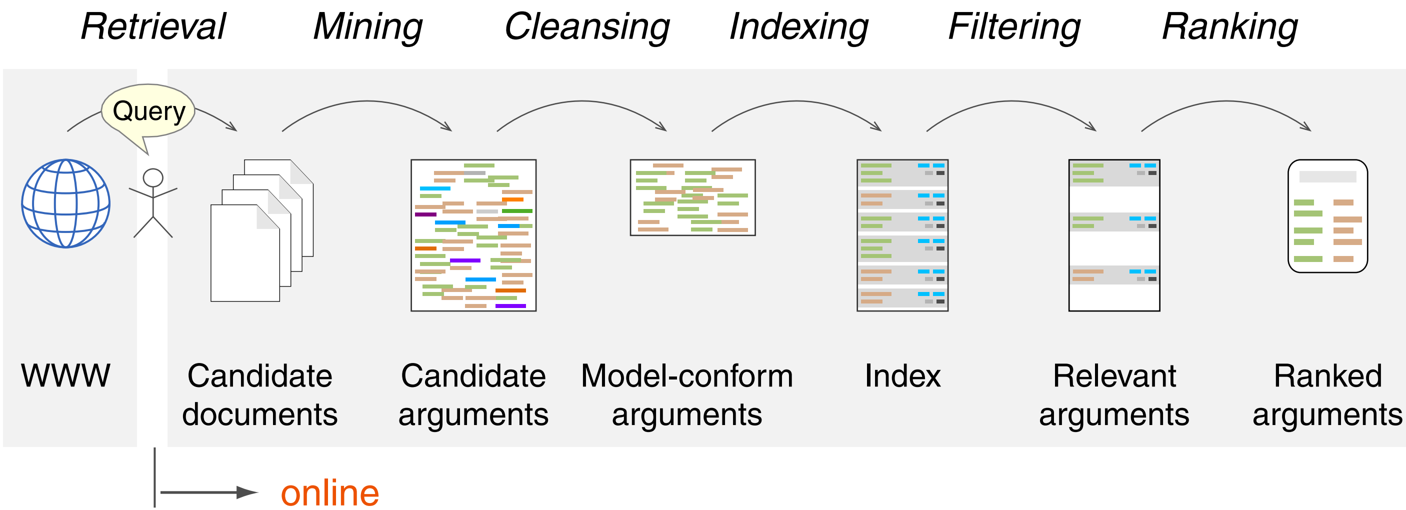
→ Effectiveness profile: high precision, high recall on topic

→ Stance balance: guaranteed

→ Efficiency: high

Argument Search Engines

Acquisition Paradigms: (c) ArgumenText [Demo]



- Research focus: argument mining
- Supervision level: low

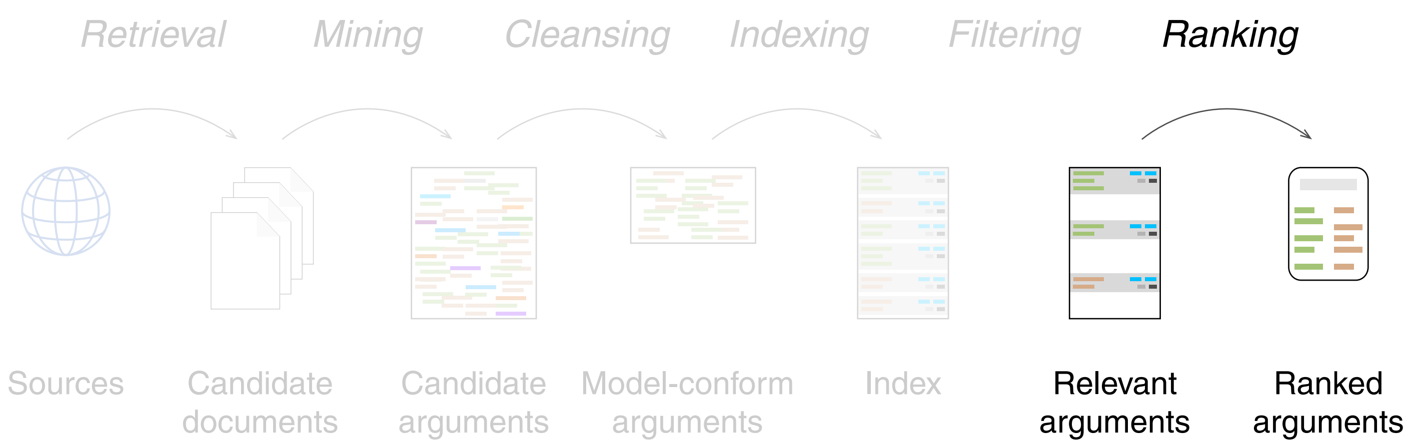
➜ Effectiveness profile: low precision, high recall

➜ Stance balance: cannot be guaranteed

➜ Efficiency: low

Argument Search Engines

Ranking Paradigms in IR



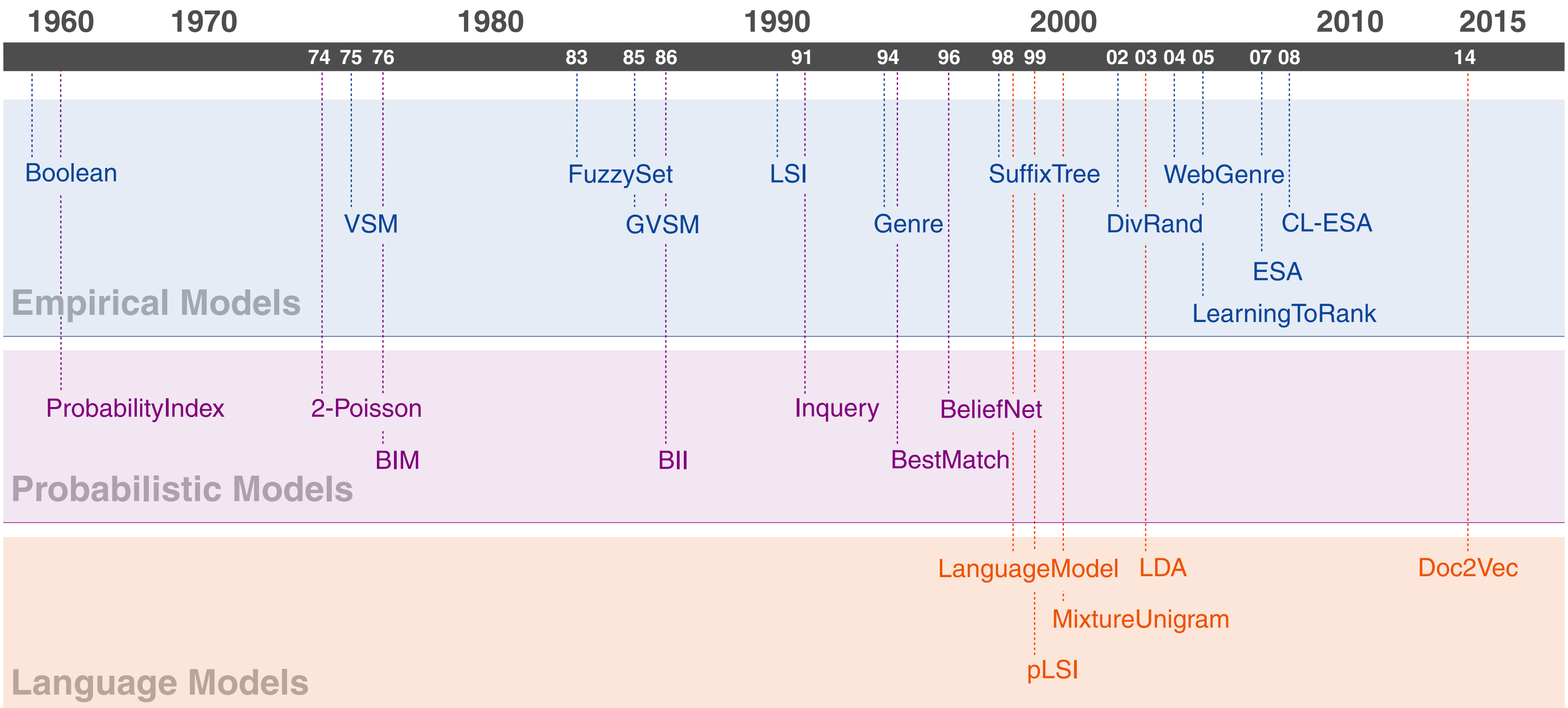
Designing a ranking algorithm:

- Analyze conclusions, premises, or both?
- Use fulltext or elite terms only?
- Exploit metadata and sentiment?
- Analyze relations between arguments?

  . . .

Argument Search Engines

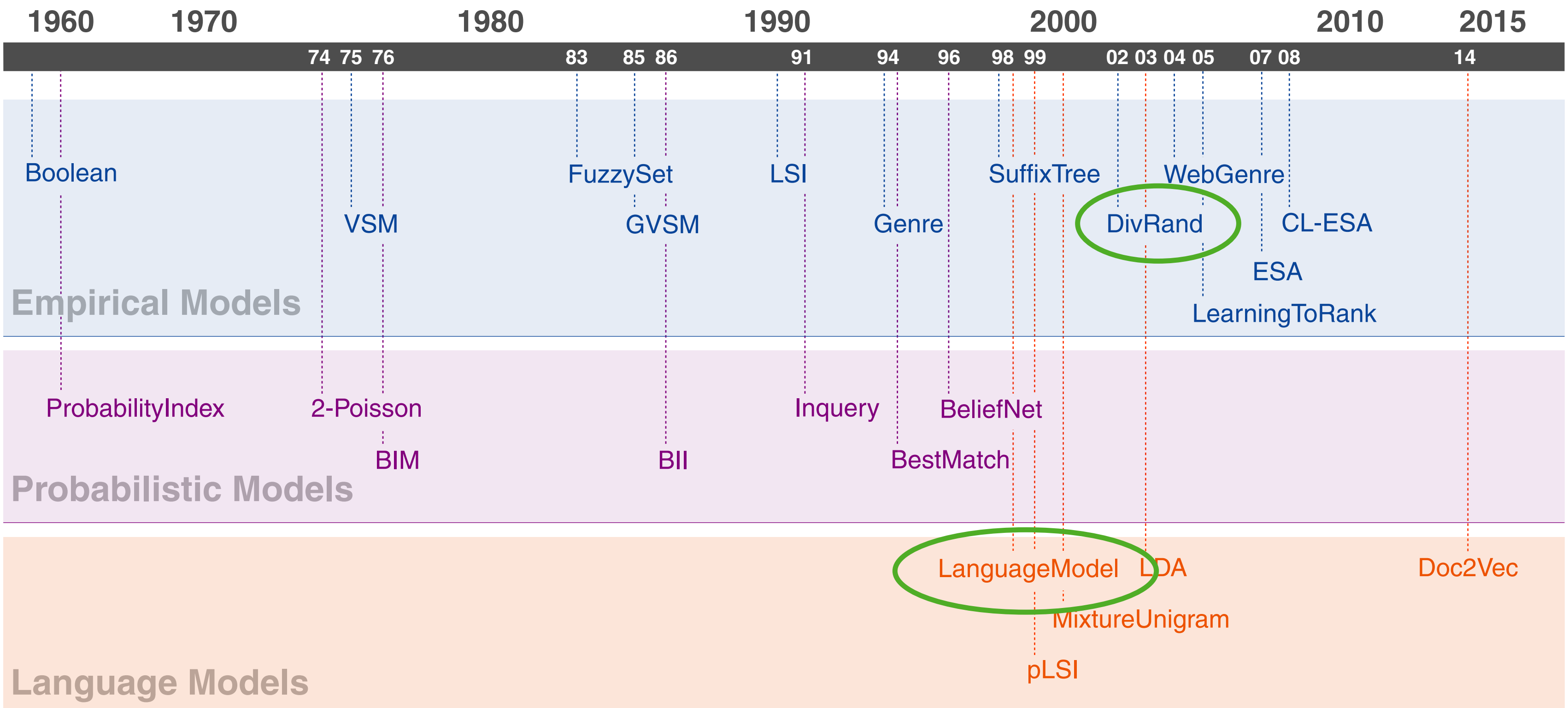
Ranking Paradigms in IR



[Stein et al. 2017]

Argument Search Engines

Ranking Paradigms in IR



- New research indicates that *Divergence from Randomness* and *Language Models* are the currently most effective retrieval models to address Π_{rank}.

  [Pottast et al. 2019]

Argument Search Evaluation

Same Side Stance Classification [sameside.webis.de]

Task: Given two arguments regarding a certain topic, decide whether or not the two arguments have the same stance.

Topic: “Gay marriage should be legalized.”

Argument 1
Marriage is a commitment to love and care for your spouse till death. This is what is heard in all wedding vows. Gays can clearly qualify for marriage according to these vows, and any definition of marriage deduced from these vows.

Argument 2
Marriage is the institution that forms and upholds for society, its values and symbols are related to procreation. To change the definition of marriage to include same-sex couples would destroy its function.

Argument Search Evaluation

Same Side Stance Classification [sameside.webis.de]

Task: Given two arguments regarding a certain topic, decide whether or not the two arguments have the same stance.

Topic: “Gay marriage should be legalized.”

Argument 1	**Argument 2**
Marriage is a commitment to love and care for your spouse till death. This is what is heard in all wedding vows. Gays can clearly qualify for marriage according to these vows, and any definition of marriage deduced from these vows.	Marriage is the institution that forms and upholds for society, its values and symbols are related to procreation. To change the definition of marriage to include same-sex couples would destroy its function.

Argument 1	**Argument 2**
Marriage is a commitment to love and care for your spouse till death. This is what is heard in all wedding vows. Gays can clearly qualify for marriage according to these vows, and any definition of marriage deduced from these vows.	Gay marriage should be legalized since denying some people the option to marry is dscrimenatory and creates a second class of citizens.

Argument Search Evaluation

Same Side Stance Classification [sameside.webis.de]

Task: Given two arguments regarding a certain topic, decide whether or not the two arguments have the same stance.

Topic: “Gay marriage should be legalized.”

Argument 1
Marriage is a commitment to love and care for your spouse till death. This is what is heard in all wedding vows. Gays can clearly qualify for marriage according to these vows, and any definition of marriage deduced from these vows.

Argument 2
Marriage is the institution that forms and upholds for society, its values and symbols are related to procreation. To change the definition of marriage to include same-sex couples would destroy its function.

◯≠◯ different side

Argument 1
Marriage is a commitment to love and care for your spouse till death. This is what is heard in all wedding vows. Gays can clearly qualify for marriage according to these vows, and any definition of marriage deduced from these vows.

Argument 2
Gay marriage should be legalized since denying some people the option to marry is dscrimenatory and creates a second class of citizens.

◯=◯ same side

Argument Search Evaluation

Same Side Stance Classification: Task Rationale

Same side classification needs not to distinguish topic-specific pro- / con-vocabulary.

→ "Only" argument similarity within a stance needs to be assessed.

→ Same side classification may be solved in a topic-agnostic fashion.

Applications:

- measure the bias strength within argumentation
- structure a discussion
- find out who or what is challenging me in a discussion
- filter wrongly labeled stances in a large argument corpus
- ...

Argument Search Evaluation

Same Side Stance Classification: Tasks Details

Two topics (domains):

1. Should gay marriage be legalized?
2. Should abortion be legalized?

Within domain setting:

Training. Instances from both domains.
Test. Instances from both domains.

Cross domain setting:

Training. Instances from abortion.
Test. Instances from gay marriage.

Argument Search Evaluation

Same Side Stance Classification: Tasks Details

Two topics (domains):

1. Should gay marriage be legalized?
2. Should abortion be legalized?

Within domain setting:

Training.	Instances from both domains.
Test.	Instances from both domains.

Cross domain setting:

Training.	Instances from abortion.
Test.	Instances from gay marriage.

Form of an instance:

1. Name of the topic (domain) d.
2. Argument 1 from $\mathbf{A}_d$.
3. Argument 2 from $\mathbf{A}_d$.
4. One of $\{ \bigcirc = \bigcirc,\ \bigcirc \neq \bigcirc \}$.

Timeline:

8.6. 2019:	Training data online.
14.6. 2019:	Submission open.
21.7. 2019:	Submission closed.
1.8. 2019:	6th ArgMining workshop.

Argument Search Evaluation

Same Side Stance Classification: Results "Within Domain"

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
...									

Argument Search Evaluation

Same Side Stance Classification: Results “Within Domain”

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
...									

Argument Search Evaluation

Same Side Stance Classification: Results "Within Domain"

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
...									

Argument Search Evaluation

Same Side Stance Classification: Results "Within Domain"

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
...									

Argument Search Evaluation

Same Side Stance Classification: Results "Within Domain"

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
...									

Trier University. BERT (large, uncased, sequence length 512), tuning for 3 epochs.

Argument Search Evaluation

Same Side Stance Classification: Results "Within Domain"

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
. . .									

Leipzig University. BERT (uncased, sequence length 512, tuning for 5 epochs), loss function: sigmoid_binary_crossentropy.

Argument Search Evaluation

Same Side Stance Classification: Results “Within Domain”

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
. . .									

IBM Research. Two BERT models fine-tuned in cascade starting from the vanilla BERT model.

Argument Search Evaluation

Same Side Stance Classification: Results “Within Domain”

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
. . .									

TU Darmstadt. Microsoft’s Multi-Task Deep Neural Network mt-dnn. Basis for the mt-dnn is BERT (large). No hyper-parameter tuning, 4 epochs.

Argument Search Evaluation

Same Side Stance Classification: Results “Within Domain”

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
...									

Düsseldorf University. Manhattan LSTM – a siamese network – which measures the similarity of both arguments. Document embeddings via BERT (base, uncased, not fine-tuned, sequence length 512 tokens).

Argument Search Evaluation

Same Side Stance Classification: Results "Within Domain"

	Gay marriage			Abortion			All		
Team	Pre	Rec	Acc	Pre	Rec	Acc	Pre	Rec	Acc
Trier University	0.90	0.73	0.83	0.79	0.59	0.71	**0.85**	0.66	0.77
Leipzig University	0.80	0.78	0.79	0.78	0.68	0.75	0.79	0.73	0.77
IBM Research	0.73	0.63	0.70	0.64	0.54	0.62	0.69	0.59	0.66
TU Darmstadt	0.74	0.56	0.68	0.63	0.48	0.60	0.68	0.52	0.64
Düsseldorf University	0.76	0.35	0.62	0.65	0.32	0.57	0.70	0.33	0.60
LMU	0.53	1.00	0.55	0.53	1.00	0.55	0.53	1.00	0.55
...									

LMU. Bert (base). Arguments organized as graph: edges are weighted with the confidence that arguments agree and confidence that they disagree. If known from training set that the arguments agree or disagree the confidence is 0 and 1 or 1 and 0 accordingly.

Argument Search Evaluation

Same Side Stance Classification: Results "Cross Domain"

	Gay marriage (large)			Gay marriage (small)		
Team	Pre	Rec	Acc	Pre	Rec	Acc
LMU	0.67	0.53	0.63	0.78	0.61	0.72
TU Darmstadt	0.64	0.59	0.63	0.71	0.63	0.68
IBM Research	0.62	0.49	0.60	0.74	0.43	0.64
Paderborn University	0.60	0.38	0.56	0.79	0.33	0.62
Trier University	0.69	0.16	0.54	**1.00**	0.20	0.60
Düsseldorf University	0.72	0.53	0.66	0.68	0.37	0.60
...						

Argument Search Evaluation

Same Side Stance Classification: Results “Cross Domain”

	Gay marriage (large)			Gay marriage (small)		
Team	Pre	Rec	Acc	Pre	Rec	Acc
LMU	0.67	0.53	0.63	0.78	0.61	0.72
TU Darmstadt	0.64	0.59	0.63	0.71	0.63	0.68
IBM Research	0.62	0.49	0.60	0.74	0.43	0.64
Paderborn University	0.60	0.38	0.56	0.79	0.33	0.62
Trier University	0.69	0.16	0.54	**1.00**	0.20	0.60
Düsseldorf University	0.72	0.53	0.66	0.68	0.37	0.60
...						

Argument Search Evaluation

Same Side Stance Classification: Results "Cross Domain"

	Gay marriage (large)			Gay marriage (small)		
Team	Pre	Rec	Acc	Pre	Rec	Acc
LMU	0.67	0.53	0.63	0.78	0.61	0.72
TU Darmstadt	0.64	0.59	0.63	0.71	0.63	0.68
IBM Research	0.62	0.49	0.60	0.74	0.43	0.64
Paderborn University	0.60	0.38	0.56	0.79	0.33	0.62
Trier University	0.69	0.16	0.54	**1.00**	0.20	0.60
Düsseldorf University	0.72	0.53	0.66	0.68	0.37	0.60
...						

Argument Search Evaluation

Same Side Stance Classification: Results “Cross Domain”

	Gay marriage (large)			Gay marriage (small)		
Team	Pre	Rec	Acc	Pre	Rec	Acc
LMU	0.67	0.53	0.63	0.78	0.61	0.72
TU Darmstadt	0.64	0.59	0.63	0.71	0.63	0.68
IBM Research	0.62	0.49	0.60	0.74	0.43	0.64
Paderborn University	0.60	0.38	0.56	0.79	0.33	0.62
Trier University	0.69	0.16	0.54	**1.00**	0.20	0.60
Düsseldorf University	0.72	0.53	0.66	0.68	0.37	0.60
. . .						

Most of the submitted classifiers are robust regarding:

- imbalances between domain proportions in training and test
- imbalances between domain proportions within test
- imbalances between same side / different side proportions

Argument Search Evaluation

Argument Retrieval Task @ CLEF 2020 [touche.webis.de]

Argument Search Evaluation

Argument Retrieval Task @ CLEF 2020 [touche.webis.de]

Task 1: Supporting argumentative conversations

- Scenario: Users search for arguments on controversial topics
- Task: Retrieve “strong” pro/con arguments on the topic
- Data: 300,000 “arguments” (short text passages)

Task 2: Answering comparative questions with arguments

- Scenario: Users face personal decisions from everyday life
- Task: Retrieve arguments for “Is X better than Y for Z?”
- Data: ClueWeb12 or ChatNoir [chatnoir.eu]

- Run submissions similar to “classical” TREC tracks
- Software submissions via TIRA [tira.io]

Argument Search Evaluation

Supporting Argumentative Conversations: Results

Team	Run	nDCG@5
Dread Pirate Roberts	1	0.808
Swordsman (Baseline)	-	0.756
Dread Pirate Roberts	2	0.755
Aragorn	1	0.684
Dread Pirate Roberts	3	0.598
Zorro	-	0.573
...		

Argument Search Evaluation

Supporting Argumentative Conversations: Results

Team	Run	nDCG@5
Dread Pirate Roberts	1	0.808
Swordsman (Baseline)	-	0.756
Dread Pirate Roberts	2	0.755
Aragorn	1	0.684
Dread Pirate Roberts	3	0.598
Zorro	-	0.573
...		

Dread Pirate Roberts. Retrieval: DirichletLM/Similarity-based. Augmentation: Language modeling.

Argument Search Evaluation

Supporting Argumentative Conversations: Results

Team	Run	nDCG@5
Dread Pirate Roberts	1	0.808
Swordsman (Baseline)	-	0.756
Dread Pirate Roberts	2	0.755
Aragorn	1	0.684
Dread Pirate Roberts	3	0.598
Zorro	-	0.573
...		

Swordsman (Baseline). Retrieval: DirichletLM.

Argument Search Evaluation

Supporting Argumentative Conversations: Results

Team	Run	nDCG@5
Dread Pirate Roberts	1	0.808
Swordsman (Baseline)	-	0.756
Dread Pirate Roberts	2	0.755
Aragorn	1	0.684
Dread Pirate Roberts	3	0.598
Zorro	-	0.573
. . .		

Dread Pirate Roberts. Retrieval: DirichletLM/Similarity-based. Augmentation: Language modeling.

Argument Search Evaluation

Supporting Argumentative Conversations: Results

Team	Run	nDCG@5
Dread Pirate Roberts	1	0.808
Swordsman (Baseline)	-	0.756
Dread Pirate Roberts	2	0.755
Aragorn	1	0.684
Dread Pirate Roberts	3	0.598
Zorro	-	0.573
...		

Aragorn: Retrieval. BM25. (Re)ranking Feature: Premise prediction.

Argument Search Evaluation

Supporting Argumentative Conversations: Results

Team	Run	nDCG@5
Dread Pirate Roberts	1	0.808
Swordsman (Baseline)	-	0.756
Dread Pirate Roberts	2	0.755
Aragorn	1	0.684
Dread Pirate Roberts	3	0.598
Zorro	-	0.573
. . .		

Dread Pirate Roberts. Retrieval: DirichletLM/Similarity-based. Augmentation: Language modeling.

Argument Search Evaluation

Supporting Argumentative Conversations: Results

Team	Run	nDCG@5
Dread Pirate Roberts	1	0.808
Swordsman (Baseline)	-	0.756
Dread Pirate Roberts	2	0.755
Aragorn	1	0.684
Dread Pirate Roberts	3	0.598
Zorro	-	0.573
. . .		

Zorro: Retrieval. BM25. (Re)ranking Feature: Quality + NER.

Argument Search Evaluation

Answering Comparative Questions with Arguments: Results

Team	Run	nDCG@5
Bilbo Baggins	-	0.580
Puss in Boots (ChatNoir)	-	0.568
Inigo Montoya	-	0.567
Katana	1	0.564
Katana	2	0.553
Katana	3	0.464
. . .		

Argument Search Evaluation

Answering Comparative Questions with Arguments: Results

Team	Run	nDCG@5
Bilbo Baggins	-	0.580
Puss in Boots (ChatNoir)	-	0.568
Inigo Montoya	-	0.567
Katana	1	0.564
Katana	2	0.553
Katana	3	0.464
...		

Bilbo Baggins. Representation: Bag of words. Query processing: Named entities, comp. aspects. (Re-)Ranking features: Credibility, support.

Argument Search Evaluation

Answering Comparative Questions with Arguments: Results

Team	Run	nDCG@5
Bilbo Baggins	-	0.580
Puss in Boots (ChatNoir)	-	0.568
Inigo Montoya	-	0.567
Katana	1	0.564
Katana	2	0.553
Katana	3	0.464
. . .		

Puss in Boots (ChatNoir). Representation: Bag of words. (Re-)Ranking features: BM25F, SpamRank.

Argument Search Evaluation

Answering Comparative Questions with Arguments: Results

Team	Run	nDCG@5
Bilbo Baggins	-	0.580
Puss in Boots (ChatNoir)	-	0.568
Inigo Montoya	-	0.567
Katana	1	0.564
Katana	2	0.553
Katana	3	0.464
...		

Inigo Montoya. Representation: Bag of words. Query processing: Tokens & logic. OR. (Re-)Ranking features: Argum. units (TARGER).

Argument Search Evaluation

Answering Comparative Questions with Arguments: Results

Team	Run	nDCG@5
Bilbo Baggins	-	0.580
Puss in Boots (ChatNoir)	-	0.568
Inigo Montoya	-	0.567
Katana	1	0.564
Katana	2	0.553
Katana	3	0.464
. . .		

Katana. Representation:Diff. language models. Query processing: Diff. language models. (Re-)Ranking features: Comparativeness score.

Argument Search Evaluation

Answering Comparative Questions with Arguments: Results

Team	Run	nDCG@5
Bilbo Baggins	-	0.580
Puss in Boots (ChatNoir)	-	0.568
Inigo Montoya	-	0.567
Katana	1	0.564
Katana	2	0.553
Katana	3	0.464
. . .		

Katana. Representation:Diff. language models. Query processing: Diff. language models. (Re-)Ranking features: Comparativeness score.

Argument Search Evaluation

Answering Comparative Questions with Arguments: Results

Team	Run	nDCG@5
Bilbo Baggins	-	0.580
Puss in Boots (ChatNoir)	-	0.568
Inigo Montoya	-	0.567
Katana	1	0.564
Katana	2	0.553
Katana	3	0.464
. . .		

Katana. Representation:Diff. language models. Query processing: Diff. language models. (Re-)Ranking features: Comparativeness score.

Summary

3.1 Argument Retrieval Problems

basic argument model, relevant retrieval problems

3.2 Argument Ranking

topic-agnostic solution for Π_{rank} and Π_{counter}

3.3 Argument Search Engines

acquisition paradigm trades between precision, recall, and topicality

3.4 Shared Tasks

build community, acquire knowledge, improve benchmarks

Summary

3.1 Argument Retrieval Problems

basic argument model, relevant retrieval problems

3.2 Argument Ranking

topic-agnostic solution for Π_{rank} and Π_{counter}

3.3 Argument Search Engines

acquisition paradigm trades between precision, recall, and topicality

3.4 Shared Tasks

build community, acquire knowledge, improve benchmarks

Summary

3.1 Argument Retrieval Problems

basic argument model, relevant retrieval problems

3.2 Argument Ranking

topic-agnostic solution for Π_{rank} and Π_{counter}

3.3 Argument Search Engines

acquisition paradigm trades between precision, recall, and topicality

3.4 Shared Tasks

build community, acquire knowledge, improve benchmarks

Summary

3.1 Argument Retrieval Problems

basic argument model, relevant retrieval problems

3.2 Argument Ranking

topic-agnostic solution for Π_{rank} and Π_{counter}

3.3 Argument Search Engines

acquisition paradigm trades between precision, recall, and topicality

3.4 Shared Tasks

build community, acquire knowledge, improve benchmarks

Summary

3.1 Argument Retrieval Problems

basic argument model, relevant retrieval problems

3.2 Argument Ranking

topic-agnostic solution for Π_{rank} and Π_{counter}

3.3 Argument Search Engines

acquisition paradigm trades between precision, recall, and topicality

3.4 Shared Tasks

build community, acquire knowledge, improve benchmarks

Summary

3.1 Argument Retrieval Problems

basic argument model, relevant retrieval problems

3.2 Argument Ranking

topic-agnostic solution for Π_{rank} and Π_{counter}

3.3 Argument Search Engines

acquisition paradigm trades between precision, recall, and topicality

3.4 Shared Tasks

build community, acquire knowledge, improve benchmarks