Axiomatic Result Re-Ranking

October 19, 2017

Matthias Hagen
Bauhaus-Universität Weimar

matthias.hagen@uni-weimar.de

@matthias_hagen

Joint work with Michael Völske, Steve Göring, and Benno Stein

Axiomatic Result Re-Ranking

TL;DL

- Axiomatic IR has suggested constraints a retrieval model should fulfill
- So far, mostly theoretical analyses and some slight tweaks for some models

- We incorporate the axioms directly into the retrieval process via re-ranking
- Large-scale study on the ClueWeb corpora to show feasibility

A Brief Tour of Axiomatic IR

A Brief Tour of Axiomatic IR

Observations

- A lot of “basis” retrieval models perform similarly well, although motivated rather differently (BM25, PL2, Query Likelihood, . . .).
- Often minor variations of such models fail in some way; why?

Axiomatic IR Answer

- The models share beneficial properties, independent of their motivation.

Axiomatic IR Research

- Identify and formalize beneficial properties as constraints / axioms.

A Brief Tour of Axiomatic IR

Axioms

Successful retrieval functions share similar properties

Example

$$BM25(Q,D) = \sum_{i=1}^{n} IDF(q_i) \cdot \frac{TF(q_i,D) \cdot (k_1+1)}{TF(q_i,D) \; + \; k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)}$$

A Brief Tour of Axiomatic IR

Axioms

Successful retrieval functions share similar properties

- TF weighting

Example

$$BM25(Q,D) = \sum_{i=1}^{n} IDF(q_i) \cdot \frac{TF(q_i,D) \cdot (k_1+1)}{TF(q_i,D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)}$$

A Brief Tour of Axiomatic IR

Axioms

Successful retrieval functions share similar properties

- TF weighting
- IDF weighting

Example

$$BM25(Q,D) = \sum_{i=1}^{n} IDF(q_i) \cdot \frac{TF(q_i,D) \cdot (k_1+1)}{TF(q_i,D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)}$$

A Brief Tour of Axiomatic IR

Axioms

Successful retrieval functions share similar properties

- TF weighting
- IDF weighting
- Length normalization

Example

$$BM25(Q,D) = \sum_{i=1}^{n} IDF(q_i) \cdot \frac{TF(q_i,D)\cdot(k_1+1)}{TF(q_i,D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)}$$

A Brief Tour of Axiomatic IR

Axioms

Successful retrieval functions share similar properties

- TF weighting
- IDF weighting
- Length normalization

Example

$$BM25(Q,D) = \sum_{i=1}^{n} IDF(q_i) \cdot \frac{TF(q_i,D)\cdot(k_1+1)}{TF(q_i,D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)}$$

“Axioms” formally capture such properties / constraints and their application.

A Brief Tour of Axiomatic IR

Axioms

Purpose	Acronyms	Source
Term frequency	TFC1–TFC3	[Fang, Tao, Zhai; SIGIR'04]
	TDC	[Fang, Tao, Zhai; SIGIR'04]
Document length	LNC1 + LNC2	[Fang, Tao, Zhai; SIGIR'04]
	TF-LNC	[Fang, Tao, Zhai; SIGIR'04]
	QLNC	[Cummins, O'Riordan; CIKM'12]
Lower bound	LB1 + LB2	[Lv, Zhai; CIKM'11]
Query aspects	REG	[Zheng, Fang; ECIR'10]
	DIV	[Gollapurdi, Sharma; WWW'09]
Semantic similarity	STMC1 + STMC2	[Fang, Zhai; SIGIR'06]
	STMC3	[Fang, Zhai; SIGIR'06]
	TSSC1 + TSSC2	[Fang, Zhai; SIGIR'06]
Term proximity	PHC + CCC	[Tao, Zhai; SIGIR'07]

A Brief Tour of Axiomatic IR

Term Frequency Constraints

TFC1	Higher scores for documents with more occurrences of query terms.
TFC2	Score increase per added term decreases for more added terms.
TFC3	Favor documents with more distinct query terms.

Length Normalization Constaints

LNC1	Penalize long documents.
LNC2	Avoid over-penalizing long documents.
TF-LNC	Regularize the interaction of TF and document length.

Lower-bounding Term Frequency Constraints

LB1	Presence-absence gap not to be closed by length normalization.
LB2	First occurrence more important than repeated occurrence.

A Brief Tour of Axiomatic IR

Term Frequency Constraints

TFC1	Higher scores for documents with more occurrences of query terms.
TFC2	Score increase per added term decreases for more added terms.
TFC3	Favor documents with more distinct query terms.

Length Normalization Constaints

LNC1	Penalize long documents.
LNC2	Avoid over-penalizing long documents.
TF-LNC	Regularize the interaction of TF and document length.

Lower-bounding Term Frequency Constraints

LB1	Presence-absence gap not to be closed by length normalization.
LB2	First occurrence more important than repeated occurrence.

A Brief Tour of Axiomatic IR

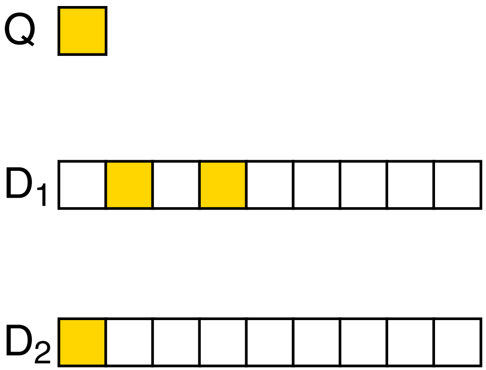
Axiom Example: TFC1

TFC1 Higher scores for documents with more occurrences of query terms.

Given:

- Single-term query $Q = \{q\}$
- Documents D_1, D_2 with $|D_1| = |D_2|$

IF $TF(q, D_1) > TF(q, D_2)$ THEN $Score(Q, D_1) > Score(Q, D_2)$



A Brief Tour of Axiomatic IR

Axiom Example: LB2

LB2 First occurrence is more important than repeated occurrence.

Given:

- Two-term query $Q = \{q_1, q_2\}$

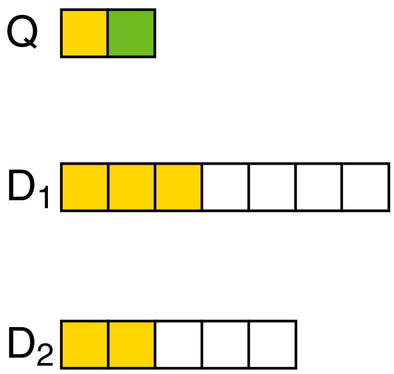
Q 

Axiom Example: LB2

LB2 First occurrence is more important than repeated occurrence.

Given:

- Two-term query $Q = \{q_1, q_2\}$
- Documents D_1, D_2 with $TF(q_1, D_i) > 0$ and $TF(q_2, D_i) = 0$



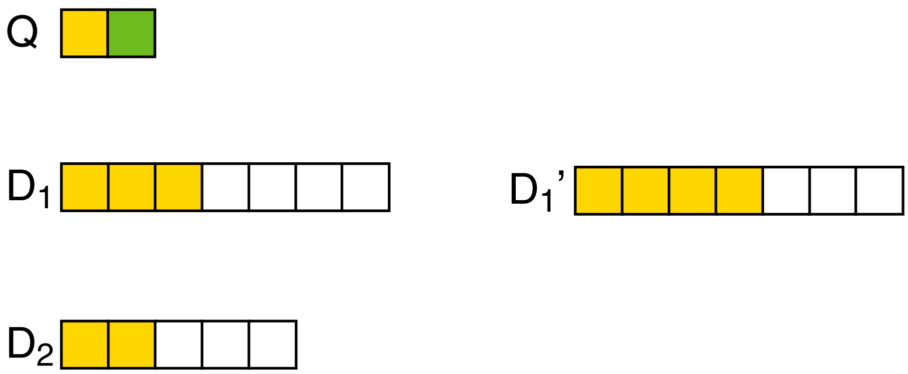
A Brief Tour of Axiomatic IR

Axiom Example: LB2

LB2 First occurrence is more important than repeated occurrence.

Given:

- Two-term query $Q = \{q_1, q_2\}$
- Documents D_1, D_2 with $TF(q_1, D_i) > 0$ and $TF(q_2, D_i) = 0$
- Document $D_1' = D_1 \cup \{q_1\} \setminus \{t_1\}$ for some $t_1 \in D_1$, $t_1 \notin Q$



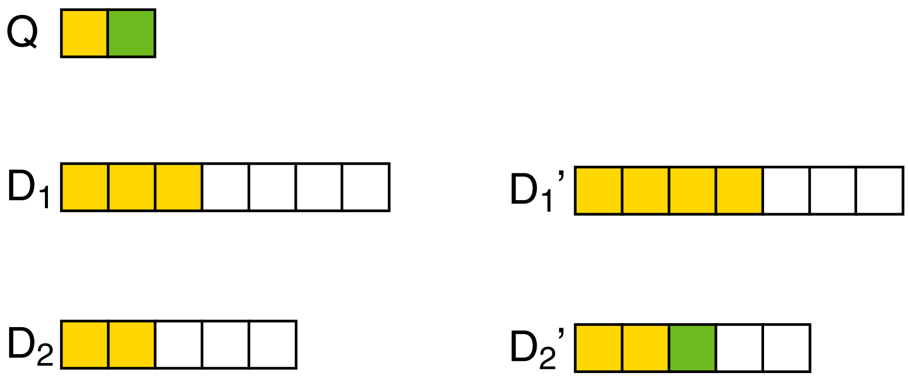
A Brief Tour of Axiomatic IR

Axiom Example: LB2

LB2 First occurrence is more important than repeated occurrence.

Given:

- Two-term query $Q = \{q_1, q_2\}$
- Documents D_1, D_2 with $TF(q_1, D_i) > 0$ and $TF(q_2, D_i) = 0$
- Document $D_1' = D_1 \cup \{q_1\} \setminus \{t_1\}$ for some $t_1 \in D_1$, $t_1 \notin Q$
- Document $D_2' = D_2 \cup \{q_2\} \setminus \{t_2\}$ for some $t_2 \in D_2$, $t_2 \notin Q$



A Brief Tour of Axiomatic IR

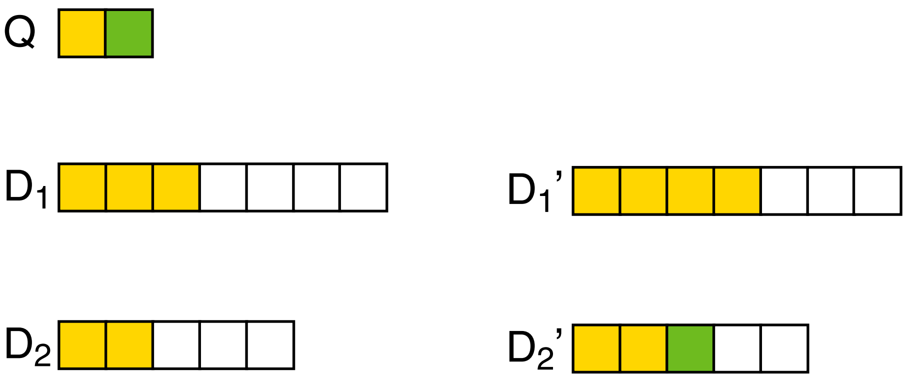
Axiom Example: LB2

LB2 First occurrence is more important than repeated occurrence.

Given:

- Two-term query $Q = \{q_1, q_2\}$
- Documents D_1, D_2 with $TF(q_1, D_i) > 0$ and $TF(q_2, D_i) = 0$
- Document $D_1' = D_1 \cup \{q_1\} \setminus \{t_1\}$ for some $t_1 \in D_1$, $t_1 \notin Q$
- Document $D_2' = D_2 \cup \{q_2\} \setminus \{t_2\}$ for some $t_2 \in D_2$, $t_2 \notin Q$

IF $Score(Q, D_1) = Score(Q, D_2)$ THEN $Score(Q, D_1') < Score(Q, D_2')$



A Brief Tour of Axiomatic IR

Axiomatic Analysis

- BM25 (no matter the parameter setting) violates the LB2 constraint
- A minor modification corrects this, for consistently better performance

  [Lv and Zhai, CIKM'11]

$$BM25\ \ (Q,D) = \sum_{i=1}^{n} IDF(q_i) \ \cdot \quad \frac{TF(q_i,D)\cdot(k_1+1)}{TF(q_i,D)+k_1\cdot\left(1-b+b\cdot\frac{|D|}{avgdl}\right)}$$

A Brief Tour of Axiomatic IR

Axiomatic Analysis

- BM25 (no matter the parameter setting) violates the LB2 constraint
- A minor modification corrects this, for consistently better performance

  [Lv and Zhai, CIKM'11]

$$BM25^{+}(Q,D) = \sum_{i=1}^{n} IDF(q_i) \cdot \left(\frac{TF(q_i,D) \cdot (k_1+1)}{TF(q_i,D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)} + \delta \right)$$

A Brief Tour of Axiomatic IR

Axiomatic Analysis

- BM25 (no matter the parameter setting) violates the LB2 constraint
- A minor modification corrects this, for consistently better performance

  [Lv and Zhai, CIKM'11]

$$BM25^{+}(Q,D) = \sum_{i=1}^{n} IDF(q_i) \cdot \left(\frac{TF(q_i,D)\cdot(k_1+1)}{TF(q_i,D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{avgdl}\right)} + \delta \right)$$

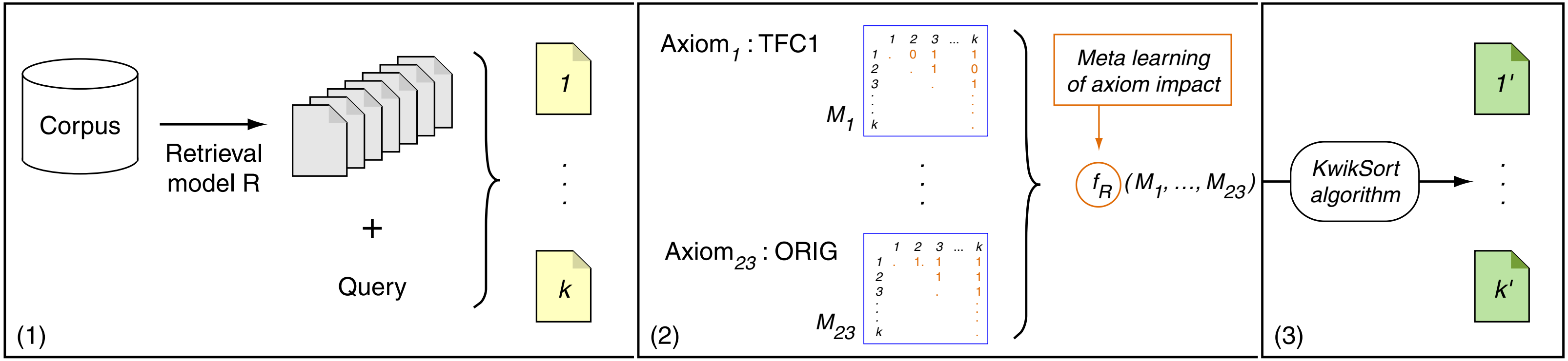
Our research question:

How can we automate such an “axiomatization” of retrieval models?

Axiomatic Result Re-Ranking

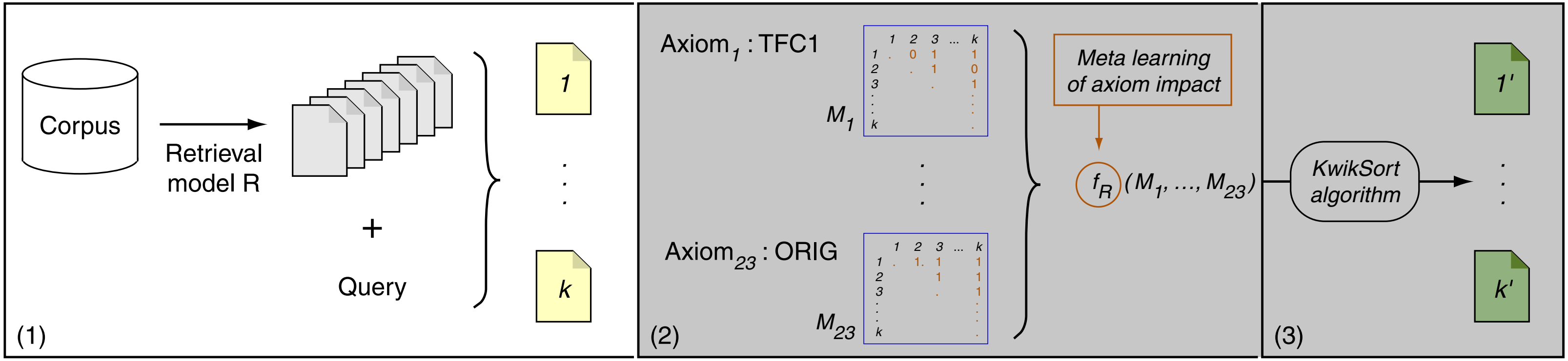
Axiomatic Result Re-Ranking

Pipeline



Axiomatic Result Re-Ranking

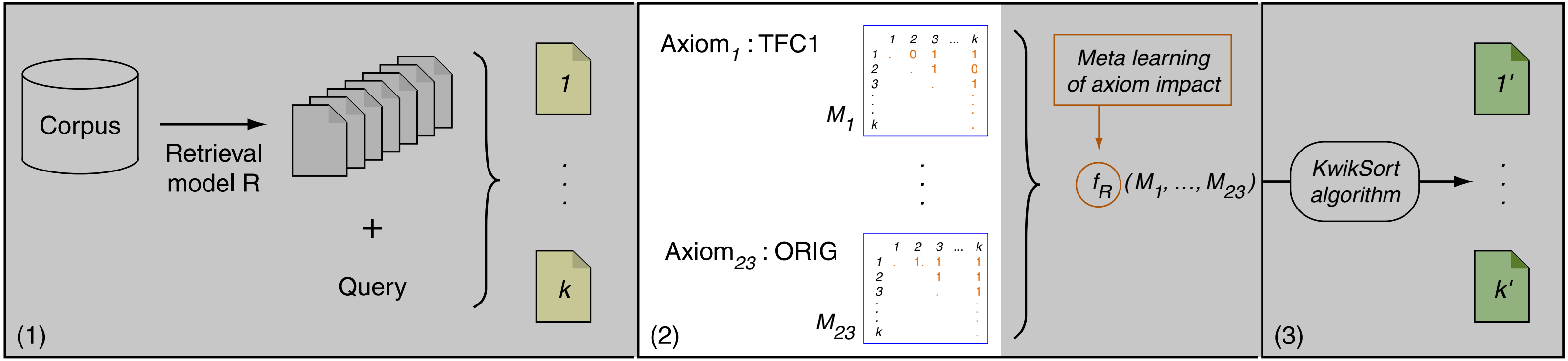
Pipeline



1. Retrieve an initial top-k result set with some basis retrieval model.

Axiomatic Result Re-Ranking

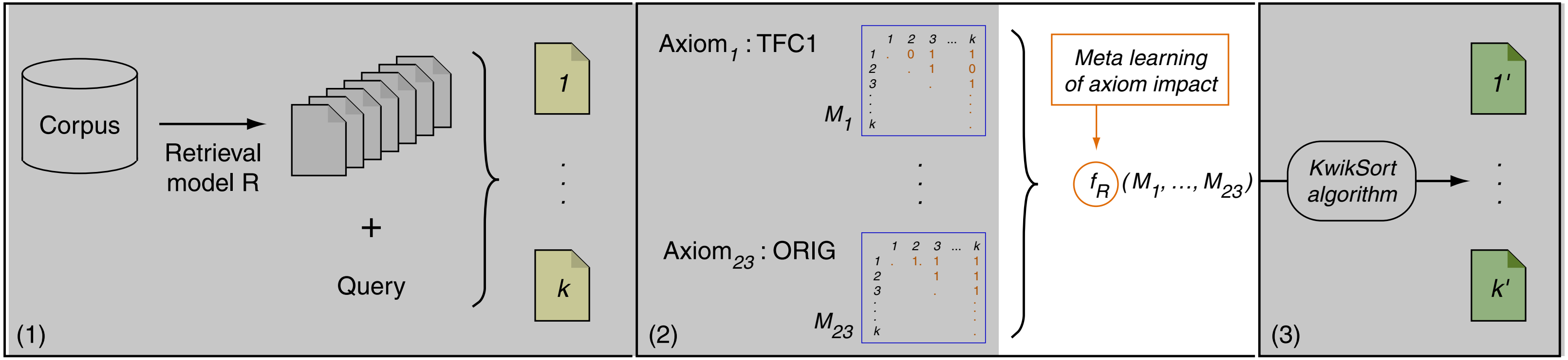
Pipeline



1. Retrieve an initial top-k result set with some basis retrieval model.

2. (a) Derive the re-ranking preferences of various axioms.

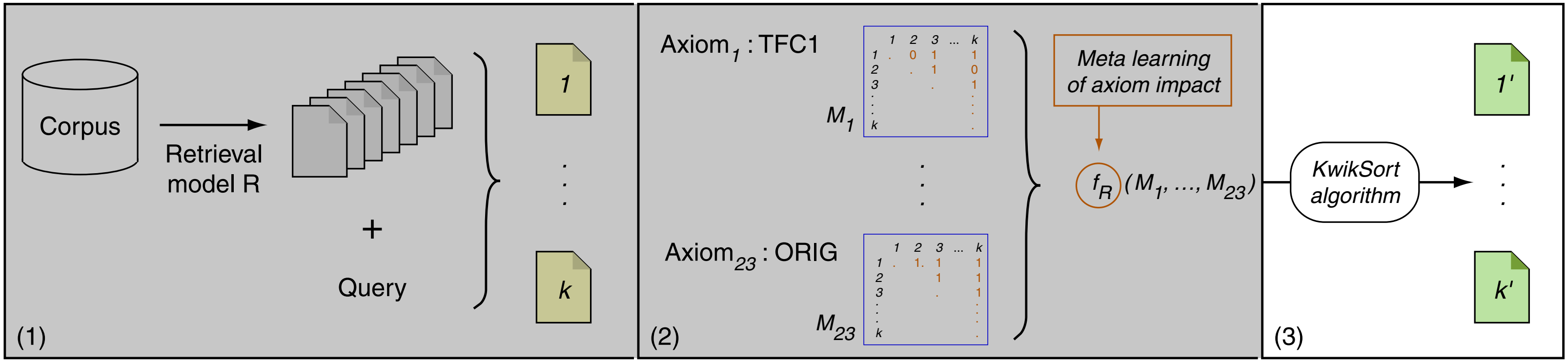
Axiomatic Result Re-Ranking
Pipeline



1. Retrieve an initial top-k result set with some basis retrieval model.
2. (a) Derive the re-ranking preferences of various axioms.
   (b) Aggregate the re-ranking preferences.

Axiomatic Result Re-Ranking

Pipeline



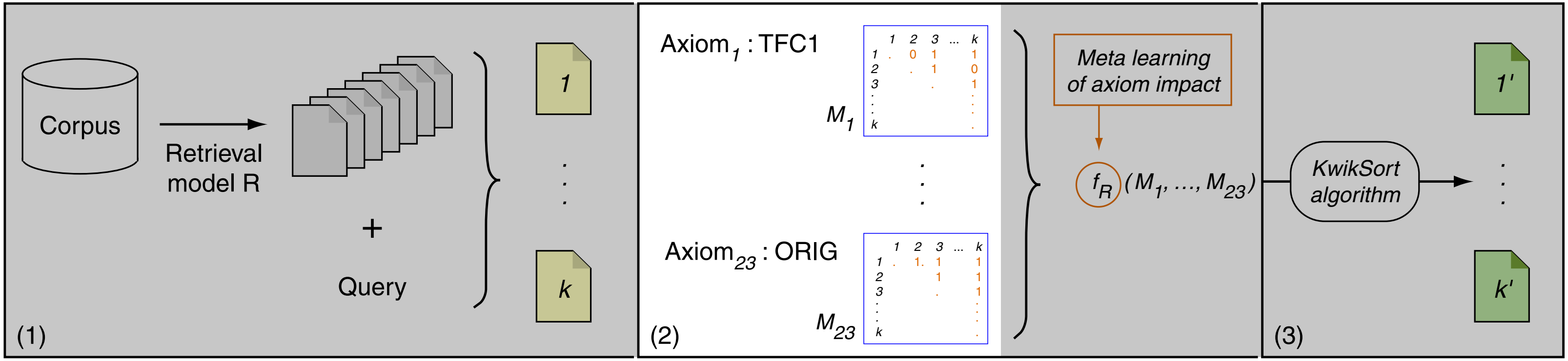
1. Retrieve an initial top-k result set with some basis retrieval model.

2. (a) Derive the re-ranking preferences of various axioms.
   (b) Aggregate the re-ranking preferences.

3. Re-rank the initial top-k result set.

Axiomatic Result Re-Ranking

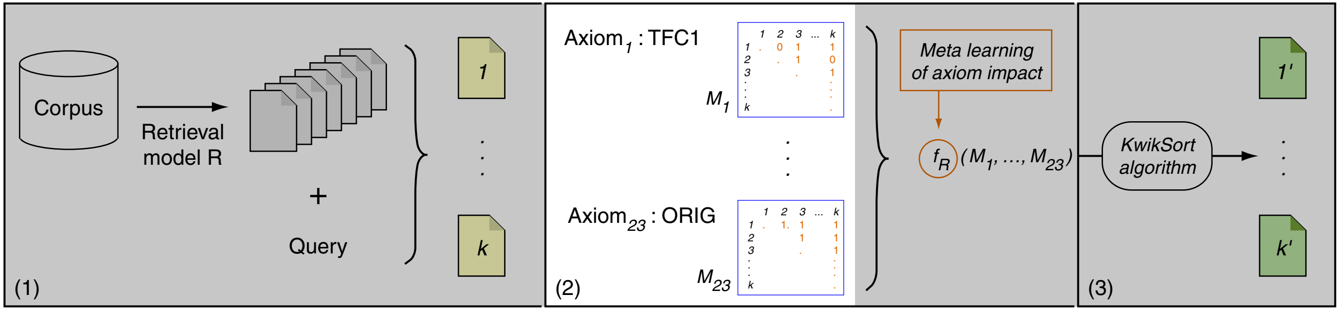
Pipeline



1. Retrieve an initial top-k result set with some basis retrieval model.
2. (a) Derive the re-ranking preferences of various axioms.
   (b) Aggregate the re-ranking preferences.
3. Re-rank the initial top-k result set.

Axiomatic Result Re-Ranking

Triple Formulation of Axioms

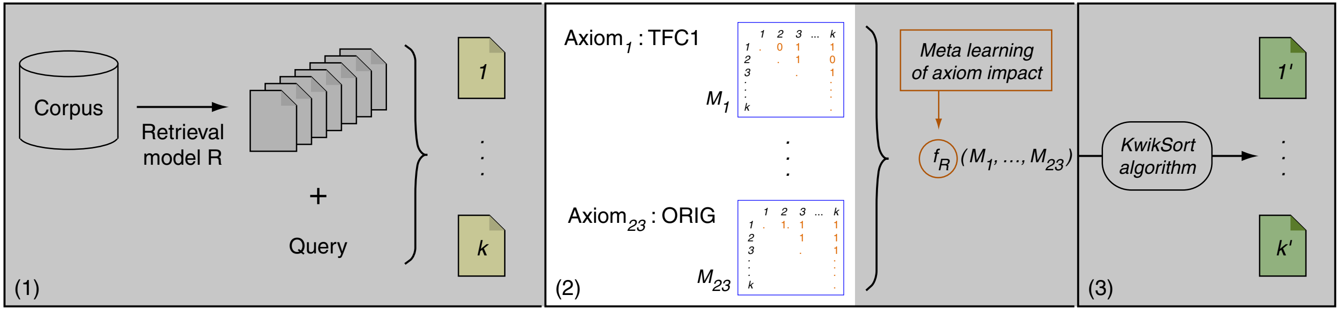


Axioms reformulated as triples

$$A = (precondition, filter, conclusion)$$

Axiomatic Result Re-Ranking

Triple Formulation of Axioms



Axioms reformulated as triples

$$A = (precondition, filter, conclusion)$$

Given axiom A and document pair D_1, D_2:

- The *precondition* indicates whether or not A can be applied to D_1, D_2
- If the *filter* condition is satisfied . . .
- . . . the *conclusion* implies a ranking preference of the form $D_1 >_A D_2$

Axiomatic Result Re-Ranking

Adapting Existing Axioms

1. Convert to triple formulation
2. Relax equality constraints and tighten inequality constraints
3. Modify conclusion to express a ranking preference

Axiomatic Result Re-Ranking

Adapting Existing Axioms

1. Convert to triple formulation
2. Relax equality constraints and tighten inequality constraints
3. Modify conclusion to express a ranking preference

Example: TFC1

Given:

- Single-term query $Q = \{q\}$
- Documents D_1, D_2 with $|D_1| = |D_2|$

IF $TF(q, D_1) > TF(q, D_2)$ THEN $Score(Q, D_1) > Score(Q, D_2)$

Axiomatic Result Re-Ranking

Adapting Existing Axioms

1. Convert to triple formulation
2. Relax equality constraints and tighten inequality constraints
3. Modify conclusion to express a ranking preference

Example: TFC1

Given:

- Single-term query $Q = \{q\}$
- Documents D_1, D_2 with $|D_1| = |D_2|$

IF $TF(q, D_1) > TF(q, D_2)$ THEN $Score(Q, D_1) > Score(Q, D_2)$

Precondition := $|D_1| \approx_{10\%} |D_2|$

Filter := $TF(q, D_1) >_{10\%} TF(q, D_2)$

Conclusion := $D_1 >_{TFC1} D_2$

Axiomatic Result Re-Ranking

Adapting Existing Axioms

Purpose	Acronyms	Adaption possible
Term frequency	TFC1–TFC3	✓
	TDC	✓
Document length	LNC1 + LNC2	✓
	TF-LNC	✓
	QLNC	✗
Lower bound	LB1 + LB2	✓
Query aspects	REG	✓
	DIV	✓
Semantic similarity	STMC1 + STMC2	✓
	STMC3	✗
	TSSC1 + TSSC2	✗
Term proximity	PHC + CCC	✗
	QPHRA	New
	PROX1–5	New
Other	ORIG	New

Axiomatic Result Re-Ranking

New Term Proximity Axioms

Given two documents D_1, D_2 and multi-term query $Q = \{q_1, q_2, \ldots, q_n\}$

Precondition: both documents contain all query terms.

Give preference to the document where:

PROX1 Query term pairs are closer together on average.

PROX2 First occurrences of query terms are earlier.

PROX3 The whole query as a phrase occurs earlier.

PROX4 The number of non-query terms in the shortest text span containing all query terms is smaller.

PROX5 The average shortest text span containing all query terms is smaller.

Axiomatic Result Re-Ranking

New Term Proximity Axioms

Given two documents D_1, D_2 and multi-term query $Q = \{q_1, q_2, \ldots, q_n\}$

PROX1 Query term pairs are closer together on average.

Q | 1 | 2 | 3

Axiomatic Result Re-Ranking

New Term Proximity Axioms

Given two documents D_1, D_2 and multi-term query $Q = \{q_1, q_2, \ldots, q_n\}$

PROX1 Query term pairs are closer together on average.

Consider query term pairs $P = \{(i, j) \mid q_i, q_j \in Q, i < j\}$

Q 1 2 3 P = { (1,2), (1,3), (2,3) }

Matthias Hagen

Axiomatic Result Re-Ranking

New Term Proximity Axioms

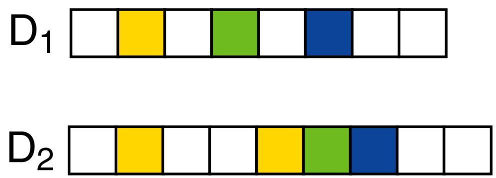
Given two documents D_1, D_2 and multi-term query $Q = \{q_1, q_2, \ldots, q_n\}$

PROX1 Query term pairs are closer together on average.

Consider query term pairs $P = \{(i,j) \mid q_i, q_j \in Q, i < j\}$

Precondition := both documents contain all query terms

P = { (1,2), (1,3), (2,3) }



Axiomatic Result Re-Ranking

New Term Proximity Axioms

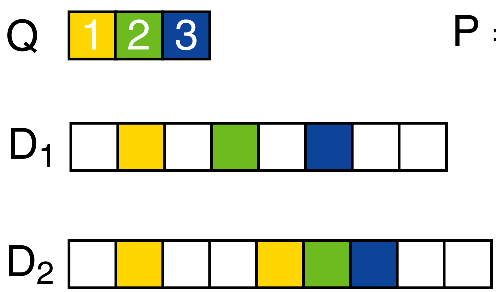
Given two documents D_1, D_2 and multi-term query $Q = \{q_1, q_2, \ldots, q_n\}$

PROX1 Query term pairs are closer together on average.

Consider query term pairs $P = \{(i,j) \mid q_i, q_j \in Q, i < j\}$ and count average number $\delta(D,i,j)$ of words between q_i, q_j in D

Precondition := both documents contain all query terms

Filter := $\pi(Q, D_1) < \pi(Q, D_2)$ where $\pi(Q,D) = \frac{1}{|P|} \sum_{(i,j) \in P} \delta(D,i,j)$



P = { (1,2), (1,3), (2,3) }

$\pi(Q, D_i)$

1/3 (1 + 3 + 1) = 5/3

Axiomatic Result Re-Ranking

New Term Proximity Axioms

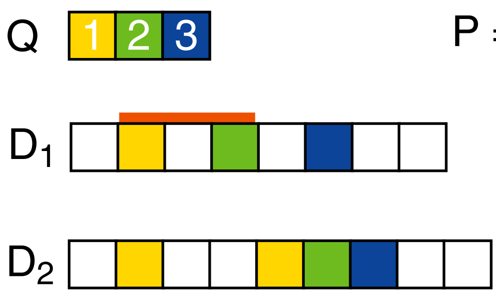
Given two documents D_1, D_2 and multi-term query $Q = \{q_1, q_2, \ldots, q_n\}$

PROX1 Query term pairs are closer together on average.

Consider query term pairs $P = \{(i,j) \mid q_i, q_j \in Q, i < j\}$ and count average number $\delta(D,i,j)$ of words between q_i, q_j in D

Precondition := both documents contain all query terms

Filter := $\pi(Q, D_1) < \pi(Q, D_2)$ where $\pi(Q, D) = \frac{1}{|P|} \sum_{(i,j) \in P} \delta(D, i, j)$



P = { (1,2), (1,3), (2,3) }

$\pi(Q,D_i)$

1/3 (1 + 3 + 1) = 5/3

Axiomatic Result Re-Ranking

New Term Proximity Axioms

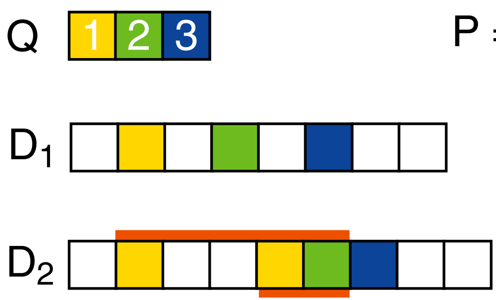
Given two documents D_1, D_2 and multi-term query $Q = \{q_1, q_2, \ldots, q_n\}$

PROX1 Query term pairs are closer together on average.

Consider query term pairs $P = \{(i,j) \mid q_i, q_j \in Q, i < j\}$ and count average number $\delta(D,i,j)$ of words between q_i, q_j in D

Precondition := both documents contain all query terms

Filter := $\pi(Q,D_1) < \pi(Q,D_2)$ where $\pi(Q,D) = \frac{1}{|P|} \sum_{(i,j)\in P} \delta(D,i,j)$



P = { (1,2), (1,3), (2,3) }

$\pi(Q, D_i)$

D_1: 1/3 (1 + 3 + 1) = 5/3

D_2: 1/3 ((3+0)/2 + (4+1)/2 + 0) = 4/3

Axiomatic Result Re-Ranking

New Term Proximity Axioms

Given two documents D_1, D_2 and multi-term query $Q = \{q_1, q_2, \ldots, q_n\}$

PROX1 Query term pairs are closer together on average.

Consider query term pairs $P = \{(i,j) \mid q_i, q_j \in Q, i < j\}$ and count average number $\delta(D,i,j)$ of words between q_i, q_j in D

Precondition := both documents contain all query terms

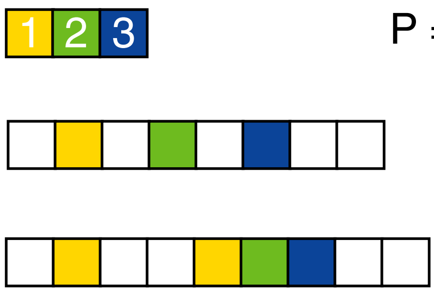
Filter := $\pi(Q, D_1) < \pi(Q, D_2)$ where $\pi(Q,D) = \frac{1}{|P|} \sum_{(i,j) \in P} \delta(D,i,j)$

Conclusion := $D_1 >_{PROX1} D_2$

Q

D_1

D_2



P = { (1,2), (1,3), (2,3) }

$\pi(Q,D_i)$

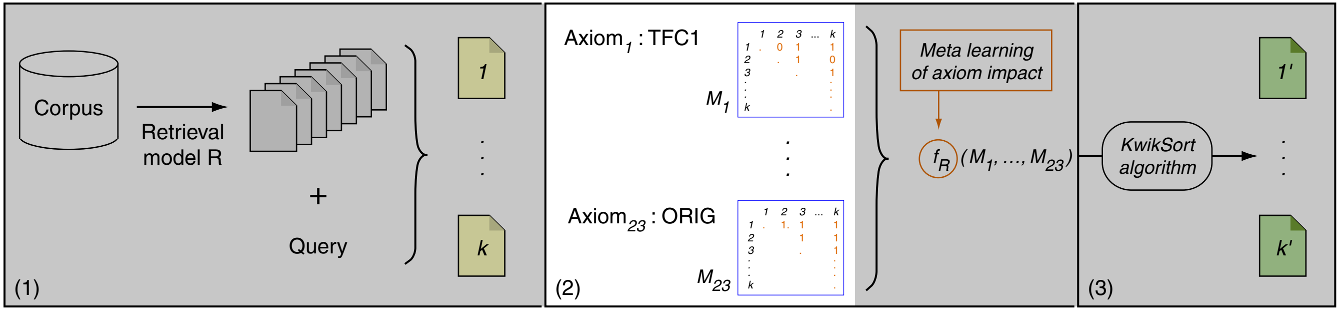
1/3 (1 + 3 + 1) = 5/3

1/3 ((3+0)/2 + (4+1)/2 + 0) = 4/3

$D_1 >_{PROX1} D_2 = 0$ $D_2 >_{PROX1} D_1 = 1$

Axiomatic Result Re-Ranking

Axiom Preference Aggregation

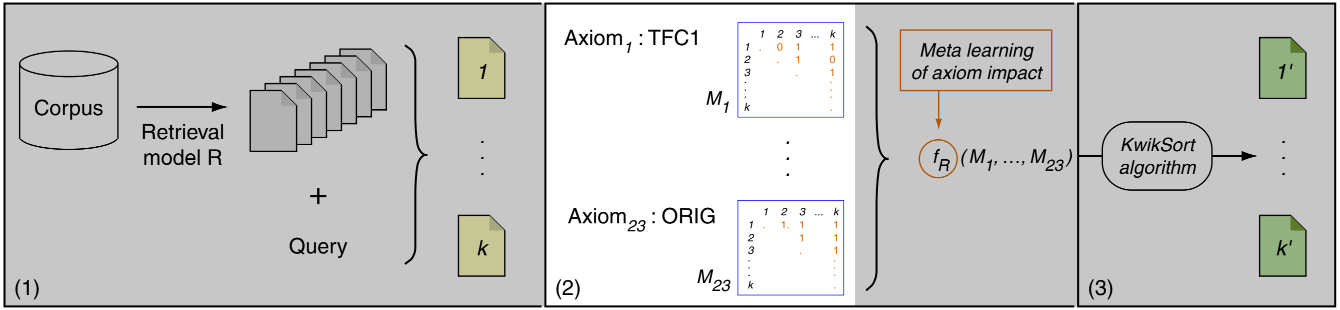


An initial result set $\{D_1, \ldots, D_k\}$ and axiom A yield a k-by-k preference matrix

$$M_A[i,j] = \begin{cases} 1 & \text{if } D_i >_A D_j, \\ 0 & \text{otherwise.} \end{cases}$$

Axiomatic Result Re-Ranking

Axiom Preference Aggregation



An initial result set $\{D_1, \ldots, D_k\}$ and axiom A yield a k-by-k preference matrix

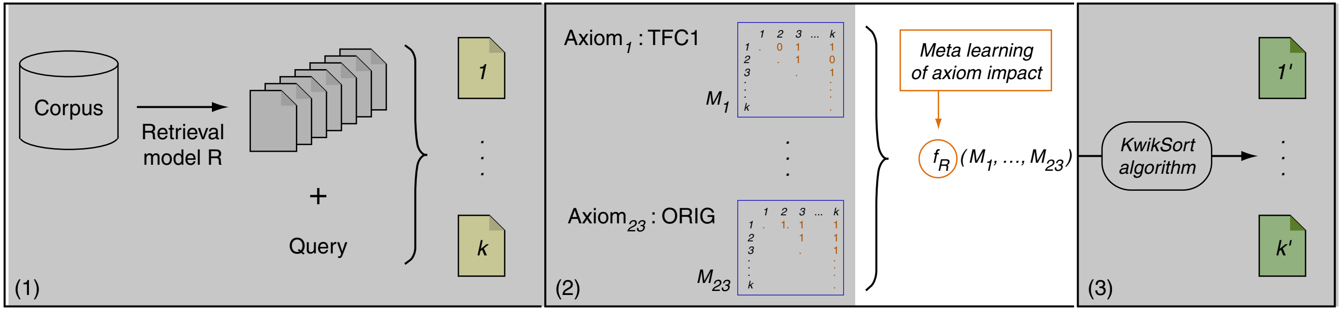
$$M_A[i,j] = \begin{cases} 1 & \text{if } D_i >_A D_j, \\ 0 & \text{otherwise.} \end{cases}$$

We include 23 axioms $\quad \rightarrow \quad$ 23 matrices:

$$A_1 : \mathrm{TFC1} \quad M_1 \quad \begin{array}{c|ccccc} & 1 & 2 & 3 & \ldots & k \\ \hline 1 & . & 0 & 1 & & 1 \\ 2 & & . & 1 & & 0 \\ 3 & & & . & & 1 \\ \vdots & & & & & \vdots \\ k & & & & & . \end{array} \qquad \cdots \qquad A_{23} : \mathrm{ORIG} \quad M_{23} \quad \begin{array}{c|ccccc} & 1 & 2 & 3 & \ldots & k \\ \hline 1 & . & 1 & 1 & & 1 \\ 2 & & . & 1 & & 1 \\ 3 & & & . & & 1 \\ \vdots & & & & & \vdots \\ k & & & & & . \end{array}$$

Axiomatic Result Re-Ranking

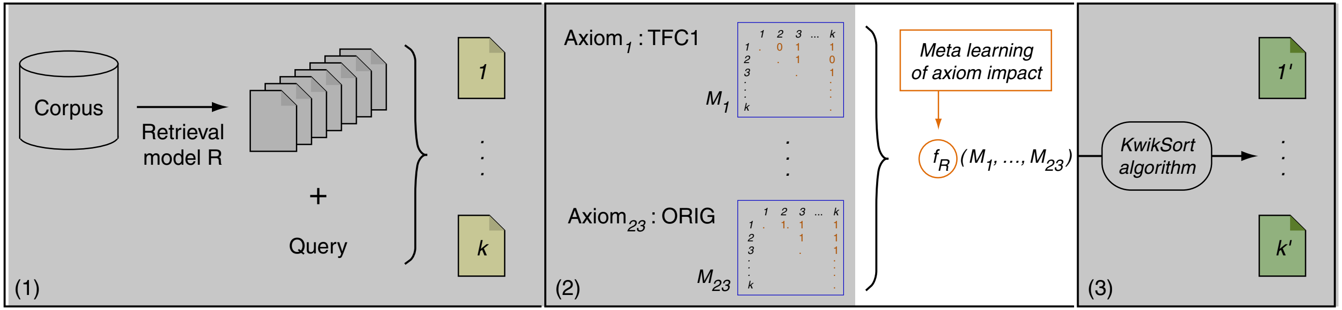
Axiom Preference Aggregation



The preferences from the 23 matrices need to be aggregated for the re-ranking.

Axiomatic Result Re-Ranking

Axiom Preference Aggregation

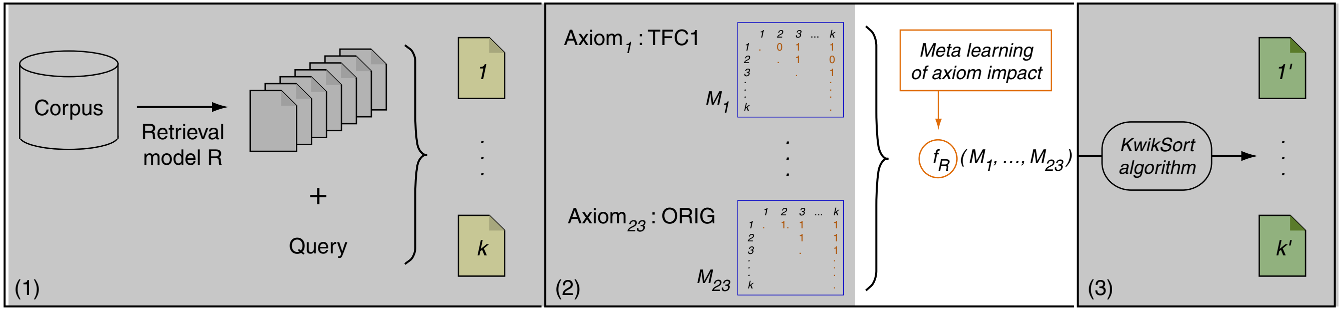


The preferences from the 23 matrices need to be aggregated for the re-ranking.

Hypothesis: Different retrieval models deviate from the “axioms” differently.
Different axioms are of different importance for different models.

Axiomatic Result Re-Ranking

Axiom Preference Aggregation



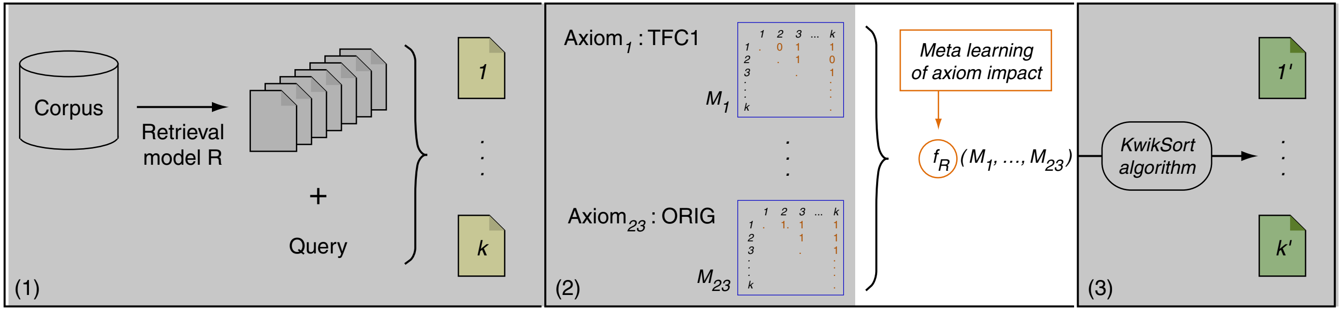
The preferences from the 23 matrices need to be aggregated for the re-ranking.

Hypothesis: Different retrieval models deviate from the “axioms” differently. Different axioms are of different importance for different models.

Approach: Learn retrieval-model-specific aggregation functions optimizing retrieval performance of the axiomatic re-ranking.

Axiomatic Result Re-Ranking

Axiom Preference Aggregation



Treat it as a classification problem

- Individual axiom preferences as predictors

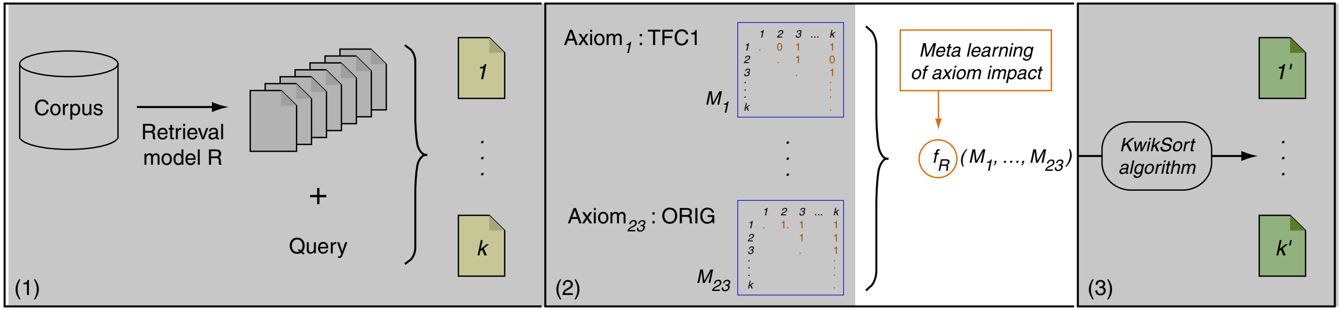
	1	2	3	...	k
1	.	0	1		1
2	0	.	1		0
3			.		1
⋮					⋮
k					.

	1	2	3	...	k
1	.	1	1		0
2	0	.	0		0
3			.		1
⋮					⋮
k					.

	1	2	3	...	k
1	.	1	1		1
2		.	1		1
3			.		1
⋮					⋮
k					.

Axiomatic Result Re-Ranking

Axiom Preference Aggregation



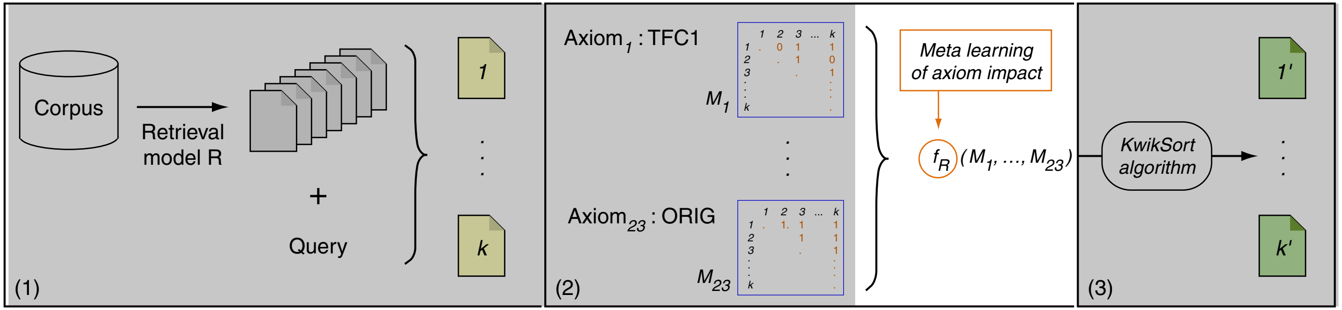
Treat it as a classification problem

- Individual axiom preferences as predictors
- Relative document relevance as response

$$
\begin{array}{cccccc}
 & 1 & 2 & 3 & \ldots & k \\
1 & . & 0 & 1 & & 1 \\
2 & 0 & . & 1 & & 0 \\
3 & & & . & & 1 \\
\vdots & & & & & \vdots \\
k & & & & & .
\end{array}
\quad
\begin{array}{cccccc}
 & 1 & 2 & 3 & \ldots & k \\
1 & . & 1 & 1 & & 0 \\
2 & 0 & . & 0 & & 0 \\
3 & & & . & & 1 \\
\vdots & & & & & \vdots \\
k & & & & & .
\end{array}
\quad
\begin{array}{cccccc}
 & 1 & 2 & 3 & \ldots & k \\
1 & . & 1 & 1 & & 1 \\
2 & & . & 1 & & 1 \\
3 & & & . & & 1 \\
\vdots & & & & & \vdots \\
k & & & & & .
\end{array}
\quad \longrightarrow \quad
\begin{array}{cccccc}
 & 1 & 2 & 3 & \ldots & k \\
1 & . & 0 & 1 & & 1 \\
2 & 1 & . & 1 & & 1 \\
3 & & & . & & 0 \\
\vdots & & & & & \vdots \\
k & & & & & .
\end{array}
$$

Axiomatic Result Re-Ranking

Axiom Preference Aggregation



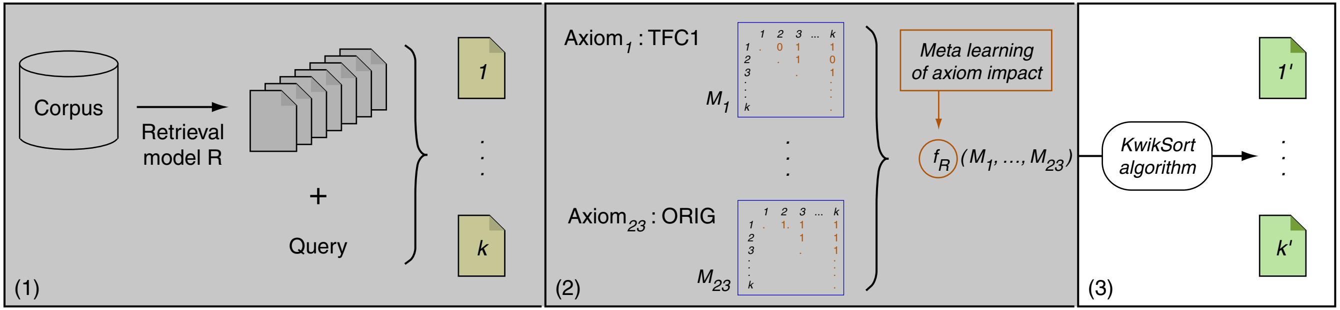
Treat it as a classification problem

- Individual axiom preferences as predictors
- Relative document relevance as response
- One training example per document pair

$$
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & \ldots & k \\
\hline
1 & . & 0 & 1 & & 1 \\
2 & 0 & . & 1 & & 0 \\
3 & & & . & & 1 \\
\vdots & & & & & \vdots \\
k & & & & & .
\end{array}
\quad
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & \ldots & k \\
\hline
1 & . & 1 & 1 & & 0 \\
2 & 0 & . & 0 & & 0 \\
3 & & & . & & 1 \\
\vdots & & & & & \vdots \\
k & & & & & .
\end{array}
\quad
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & \ldots & k \\
\hline
1 & . & 1 & 1 & & 1 \\
2 & & . & 1 & & 1 \\
3 & & & . & & 1 \\
\vdots & & & & & \vdots \\
k & & & & & .
\end{array}
\longrightarrow
\begin{array}{c|ccccc}
 & 1 & 2 & 3 & \ldots & k \\
\hline
1 & . & 0 & 1 & & 1 \\
2 & 1 & . & 1 & & 1 \\
3 & & & . & & 0 \\
\vdots & & & & & \vdots \\
k & & & & & .
\end{array}
$$

Axiomatic Result Re-Ranking

Re-ranking with Aggregated Preferences



- Aggregated matrix may contain contradictions (e.g., $M[i,j] = M[j,i]$)
- Need to solve rank-aggregation problem at this step
- We use a Kemeny rank-aggregation scheme [Kemeny; 1959]
- Approximated solution via KwikSort [Ailon, Charikar, Newman; 2008]

Experimental Evaluation

Experimental Evaluation

Axiomatic Re-Ranking at the TREC Web Track

- Take the 16 basis retrieval models of the Terrier[1] framework
- Index the ClueWeb09 using each basis retrieval model
- Retrieve top-50 results and re-rank
- 120 queries from TREC Web tracks 2009–2012 as training set
- 60 queries as test set
- Measure difference in nNDCG@10 using
  - Axiomatic re-ranking (AX)
  - Markov Random Field term dependency (MRF)
  - Both (MRF+AX)

[1] http://terrier.org

Experimental Evaluation

Average nDCG@10 on Test Set (n=60) Using Top-50 Results

Model	Basis
DPH	0.273
DFRee	0.205
In_expC2	0.205
TF_IDF	0.202
In_expB2	0.201
DFReeKLIM	0.199
BM25	0.198
InL2	0.197
BB2	0.195
DFR_BM25	0.194
LemurTF_IDF	0.187
DLH13	0.164
PL2	0.160
DLH	0.153
DirichletLM	0.139
Hiemstra_LM	0.107

Experimental Evaluation

Average nDCG@10 on Test Set (n=60) Using Top-50 Results

Model	Basis	AX
DPH	0.273	**0.291**
DFRee	0.205	**0.236**
In_expC2	0.205	**0.214**
TF_IDF	0.202	**0.228**
In_expB2	0.201	**0.202**
DFReeKLIM	0.199	**0.213**
BM25	0.198	0.188
InL2	0.197	0.197
BB2	0.195	**0.197**
DFR_BM25	0.194	**0.206**
LemurTF_IDF	0.187	**0.224**
DLH13	0.164	**0.187**
PL2	0.160	**0.213**
DLH	0.153	**0.187**
DirichletLM	0.139	**0.242**
Hiemstra_LM	0.107	**0.167**
(# higher)		14

Experimental Evaluation

Average nDCG@10 on Test Set (n=60) Using Top-50 Results

Model	Basis	AX	MRF
DPH	0.273	**0.291**	**0.307**
DFRee	0.205	**0.236**	0.230
In_expC2	0.205	**0.214**	**0.229**
TF_IDF	0.202	**0.228**	**0.239**
In_expB2	0.201	**0.202**	**0.234**
DFReeKLIM	0.199	**0.213**	**0.224**
BM25	0.198	0.188	**0.229**
InL2	0.197	0.197	**0.235**
BB2	0.195	**0.197**	**0.236**
DFR_BM25	0.194	**0.206**	**0.236**
LemurTF_IDF	0.187	**0.224**	0.221
DLH13	0.164	**0.187**	0.184
PL2	0.160	**0.213**	0.190
DLH	0.153	**0.187**	0.181
DirichletLM	0.139	**0.242**	0.192
Hiemstra_LM	0.107	**0.167**	0.161
(# higher)		14	9 (16)

Experimental Evaluation

Average nDCG@10 on Test Set (n=60) Using Top-50 Results

Model	Basis	AX	MRF	MRF+AX
DPH	0.273	**0.291**	**0.307**	**0.314**
DFRee	0.205	**0.236**	0.230	**0.245**
In_expC2	0.205	**0.214**	**0.229**	**0.238**
TF_IDF	0.202	**0.228**	**0.239**	0.200
In_expB2	0.201	**0.202**	**0.234**	**0.237**
DFReeKLIM	0.199	**0.213**	**0.224**	0.224
BM25	0.198	0.188	**0.229**	0.216
InL2	0.197	0.197	**0.235**	0.212
BB2	0.195	**0.197**	**0.236**	0.234
DFR_BM25	0.194	**0.206**	**0.236**	0.220
LemurTF_IDF	0.187	**0.224**	0.221	**0.237**
DLH13	0.164	**0.187**	0.184	**0.201**
PL2	0.160	**0.213**	0.190	**0.211**
DLH	0.153	**0.187**	0.181	**0.197**
DirichletLM	0.139	**0.242**	0.192	**0.253**
Hiemstra_LM	0.107	**0.167**	0.161	**0.163**
(# higher)		14	9 (16)	10 (15)

Experimental Evaluation

Average nDCG@10 on Test Set (n=60) Using Top-50 Results

Model	Basis	AX	MRF	MRF+AX	max
DPH	0.273	**0.291**	**0.307**	**0.314**	0.642
DFRee	0.205	**0.236**	0.230	**0.245**	0.599
In_expC2	0.205	**0.214**	**0.229**	**0.238**	0.591
TF_IDF	0.202	**0.228**	**0.239**	0.200	0.589
In_expB2	0.201	**0.202**	**0.234**	**0.237**	0.592
DFReeKLIM	0.199	**0.213**	**0.224**	0.224	0.591
BM25	0.198	0.188	**0.229**	0.216	0.587
InL2	0.197	0.197	**0.235**	0.212	0.593
BB2	0.195	**0.197**	**0.236**	0.234	0.587
DFR_BM25	0.194	**0.206**	**0.236**	0.220	0.591
LemurTF_IDF	0.187	**0.224**	0.221	**0.237**	0.576
DLH13	0.164	**0.187**	0.184	**0.201**	0.499
PL2	0.160	**0.213**	0.190	**0.211**	0.550
DLH	0.153	**0.187**	0.181	**0.197**	0.470
DirichletLM	0.139	**0.242**	0.192	**0.253**	0.564
Hiemstra_LM	0.107	**0.167**	0.161	**0.163**	0.397
(# higher)		14	9 (16)	10 (15)	16

Conclusions

Conclusions

Summary

- Axiom-based re-ranking framework for any basis retrieval model
- Directly incorporating axiomatic “thinking” in the retrieval process
- New axioms for modeling term proximity preferences

Conclusions

Summary

- Axiom-based re-ranking framework for any basis retrieval model
- Directly incorporating axiomatic “thinking” in the retrieval process
- New axioms for modeling term proximity preferences

Future Work

- Improve efficiency of the re-ranking process
- Include other axiomatic constraints (e.g., query aspects)
- Promising work in progress: Axioms for argumentation

Conclusions

Summary

- Axiom-based re-ranking framework for any basis retrieval model
- Directly incorporating axiomatic “thinking” in the retrieval process
- New axioms for modeling term proximity preferences

Future Work

- Improve efficiency of the re-ranking process
- Include other axiomatic constraints (e.g., query aspects)
- Promising work in progress: Axioms for argumentation

Thank you!