

**HYBRID QUERY
EXPANSION
MODEL FOR TEXT
AND MICROBLOG
INFORMATION
RETRIEVAL**

MORTEZA EYDIPOUR
9811634

درباره ی مقاله

- سال انتشار : ۲۰۱۸
- محل انتشار : Springer
- نویسندگان : ۵ نویسنده | دانشگاه: تونس + فرانسه

چکیده

- گسترش کوری (QE)
- HQE ?
- ما چه چیزی برای ارایه داریم ؟
- معیار کارایی و مقایسه ؟

مقدمه

- گسترش کوری :

– اضافه کردن ترم ← افزایش کارایی
روش های عملی استفاده :

Global | Local | External

روش ترکیبی چگونه؟
مشکلات موجود؟

تعاریف اولیه

• PRF

– آیا نتایج کوری برای ایجاد کوری جدیدی مرتبط مناسب است ؟

• HQE

TerGen + (PRF + WIKI)

• ترکیب دانش !!

ESAC:

ESA + WIKI

1. وابستگی آماری قوی

2. وابستگی معنایی

3. وابستگی مفهومی

تعاریف اصلی

\mathcal{C}	The <i>whole set</i> of documents which form the collection
C	A <i>set</i> of documents belonging to the collection ($C \subseteq \mathcal{C}$)
d	A <i>single</i> document of the collection ($d \in \mathcal{C}$)
V	The <i>whole set</i> of <i>distinct</i> terms of the collection \mathcal{C}
T	A <i>set</i> of terms of the collection ($T \subseteq V$)
t	A <i>single</i> term of the collection ($t \in V$)
R	An association rule
q	An original query
t_q	A term in a given query q
E_q	A query q extended

$$q = \{t_{q1}, \dots, t_{qn}\}$$

$$Supp(T) = |\{d | d \in \mathcal{C} \wedge \forall t \in T : (d, t) \in I\}|$$

$$R: T_1 \Rightarrow T_2$$

$$Supp(R) = Supp(T_1 \cup T_2)$$

$$Conf(R) = \frac{Supp(T_1 \cup T_2)}{Supp(T_1)}$$

$$minsupp \quad \quad \quad minconf$$

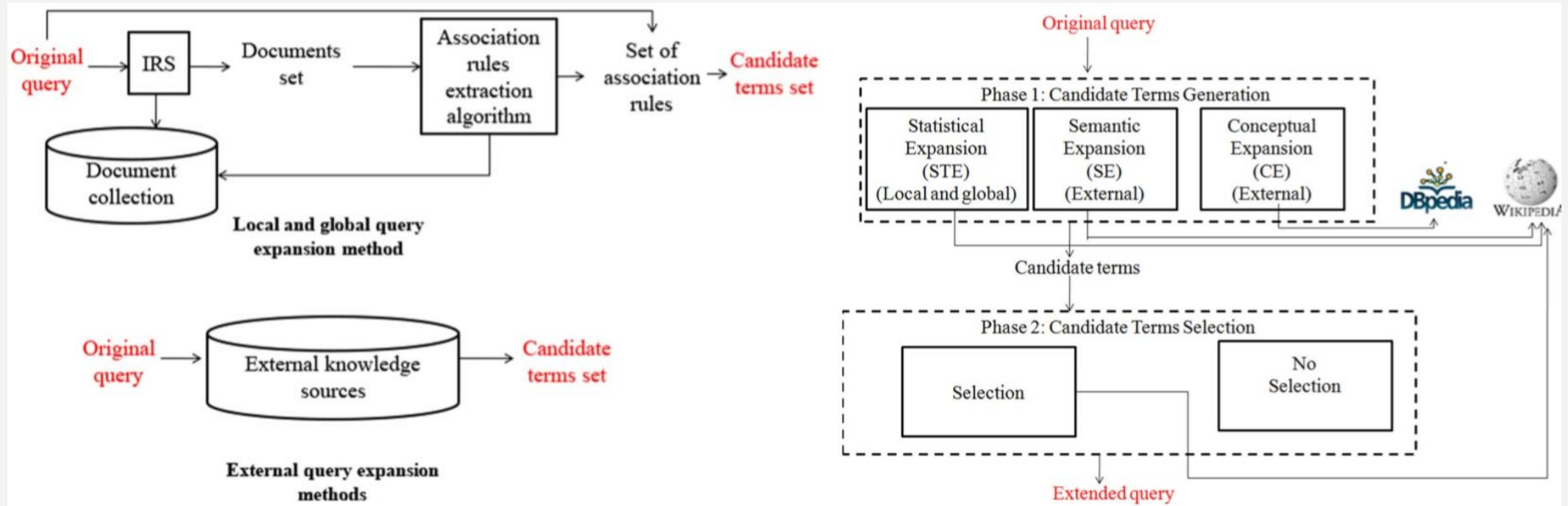
$$ESA(q, t) = \frac{\vec{q} \times \vec{t}}{\|\vec{q}\| \times \|\vec{t}\|}$$

مثال

R	Premise (T_1)	Conclusion (T_2)	$Supp(R)$	$Conf(R)$
Manufacture \Rightarrow car	Manufacture	Car	356	0.8921
Campus \Rightarrow university	Campus	University	279	0.7431
Manufacture motor \Rightarrow automobile car	Manufacture motor	Automobile car	143	0.7922

از ویکی پدیا!

HQE MODEL



سافت ترم های کاندید

$$Candidate_Set_{STE}(q) = \bigcup_{(T_1 \Rightarrow T_2) \in \mathcal{R}_C \text{ so that } T_1 \in 2^q} T_2$$

$$Candidate_Set_{SE}(q) = \bigcup_{t \in q} Def_{Semantic}(t, RS)$$

$$Candidate_Set_{CE}(q) = \bigcup_{t \in q} Concept(t, O)$$

- توسعه ی آماری
- توسعه ی معنایی
- توسعه ی مفهومی

انتخاب ترم های کاندید

$$relatedness(q, t) = score \in \mathbb{R}$$

$$E_q = q \cup \{t \in Candidate_Set(q) \mid relatedness(q, t) = score \geq \mu\}$$

$$\begin{aligned} relatedness(q, t) &= ESAC(q, t) \\ &= \begin{cases} (\alpha \times ESA(q, t) + (1 - \alpha) \times Conf_{max}(R, q, t)) & \text{if } Conf_{max}(R, q, t) \neq 0; \\ ESA(q, t), & \text{otherwise.} \end{cases} \end{aligned} \quad (16)$$

$$Conf_{max}(R, q, t) = \max_{t_q \in q, R \in \mathcal{R}_C} Conf(R(t_q, t))$$

تنظیمات HQE

Terms generation	Terms selection	
	With selection	Without selection
STE	$STE_{Selection}$	$STE_{NoSelection}$
SE	$SE_{Selection}$	$SE_{NoSelection}$
CE	$CE_{Selection}$	$CE_{NoSelection}$
ALL	$ALL_{Selection} = STE_{Selection} \cup SE_{Selection} \cup CE_{Selection}$	$ALL_{NoSelection} = STE_{NoSelection} \cup SE_{NoSelection} \cup CE_{NoSelection}$

Capital letters denote the corresponding runs in the experimental validation

بررسی عملکرد

Run	Configuration	P@5	P@10	P@30	MAP (%Chg. <i>Baseline</i> , %Chg. <i>PRF</i>)
<i>BM25</i>					
Baseline	–	0.1265	0.1327	0.1238	0.1025
PRF	–	0.1592	0.1551	0.1245	0.1145
–	STE_{Selection}	0.4000	0.3796	0.3197	0.3079 ^{†o} (200% , 168%)
–	STE _{NoSelection}	0.3551	0.3265	0.2850	0.2804 ^{†o} (173% , 145%)
–	ALL _{Selection}	0.3633	0.3429	0.2707	0.2747 ^{†o} (168% , 140%)
–	SE _{Selection}	0.3342	0.3184	0.2626	0.2589 ^{†o} (153% , 126%)
–	ALL _{NoSelection}	0.3551	0.3388	0.2553	0.2570 ^{†o} (151% , 124%)
–	CE _{Selection}	0.3224	0.3041	0.2755	0.2505 ^{†o} (144% , 118%)
–	CE _{NoSelection}	0.2408	0.2227	0.2041	0.2053 ^{†o} (100% , 79%)
–	SE _{NoSelection}	0.2367	0.2224	0.2163	0.1676 ^{†o} (63% , 46%)

بررسی عملکرد

Run	Configuration	Unigrams	Bigrams	Bigrams with 2-gaps
Best Run INEX 2013 (258)	/	0.7939	0.8908	0.8943
Median Run INEX 2013 (278)	/	0.8673	0.9540	0.9575
Worst Run INEX 2013 (269)	/	0.9981	0.9999	0.9999
/	STE _{Selection}	0.8259	0.9310	0.9302
/	SE _{Selection}	0.8172	0.9319	0.9361
/	STE _{NoSelection}	0.8279	0.9356	0.9362
/	ALL _{Selection}	0.8271	0.9374	0.9416
/	CE _{Selection}	0.8654	0.9478	0.9503
/	SE _{NoSelection}	0.8259	0.9362	0.9404
/	CE _{NoSelection}	0.8639	0.9524	0.9546