RUCIR at NTCIR-12 IMINE-2 Task

Zhicheng Dou

Renmin University of China

Joint work with Ming Yue, Sha Hu, Jinxiu Li, Xiaojie Wang, and **Ji-Rong Wen**

NTCIR-12 IMine-2 Task

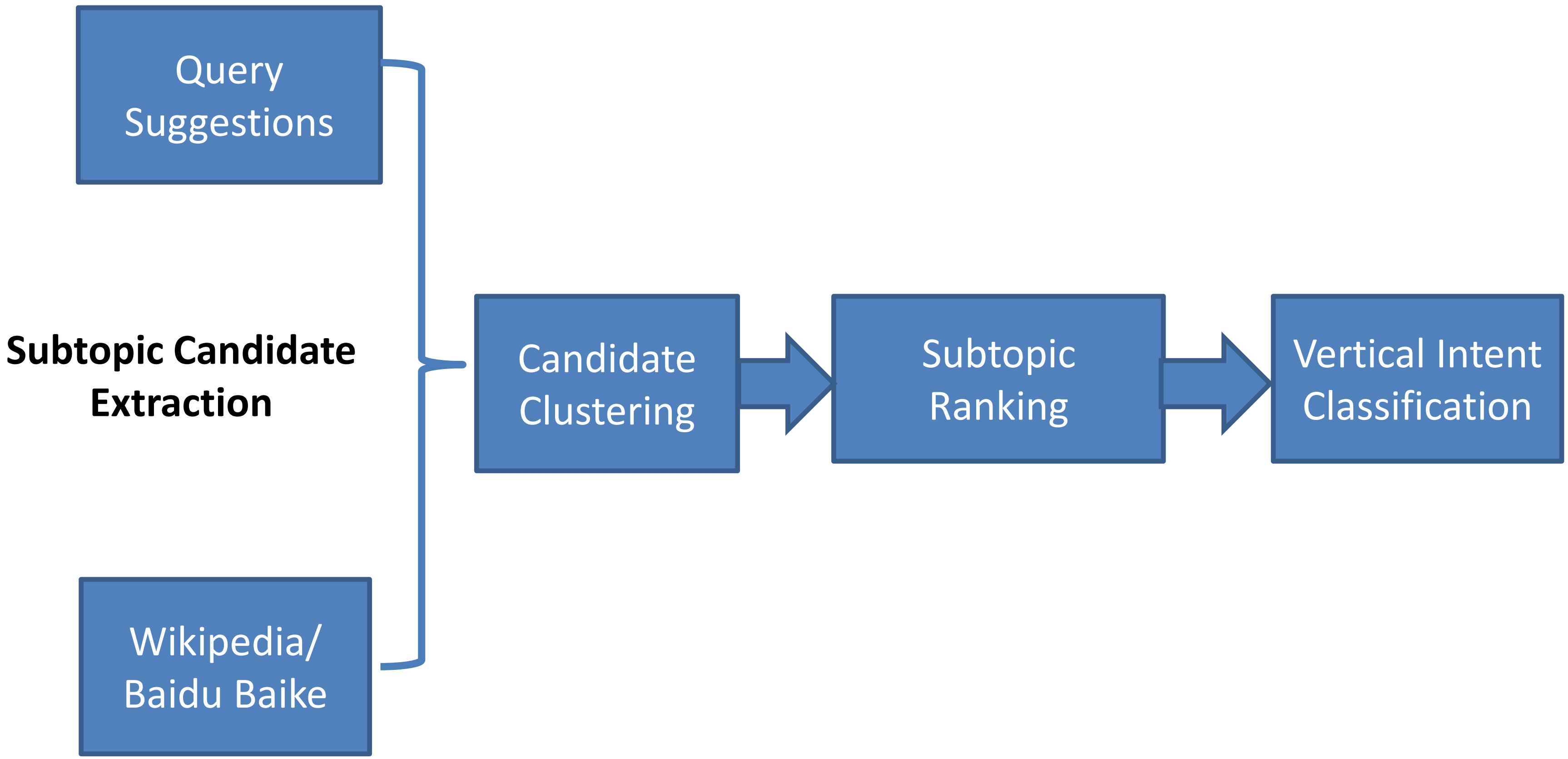
- **Goal: go toward whole-page diversification**
  - Diversify Web results together with other vertical answers
- **Subtask 1: Query Understanding**
  - Given a query, generate its subtopics, then classify their vertical intents
- **Subtask 2: Vertical Incorporating**
  - Diversify original search results together with virtual vertical documents

NTCIR-12 IMine-2 Task

- **Goal: go toward whole-page diversification**
  - Diversify Web results together with other vertical answers
- **Subtask 1: Query Understanding**
  - Given a query, generate its subtopics, then classify their vertical intents
- **Subtask 2: Vertical Incorporating**
  - Diversify original search results together with virtual vertical documents

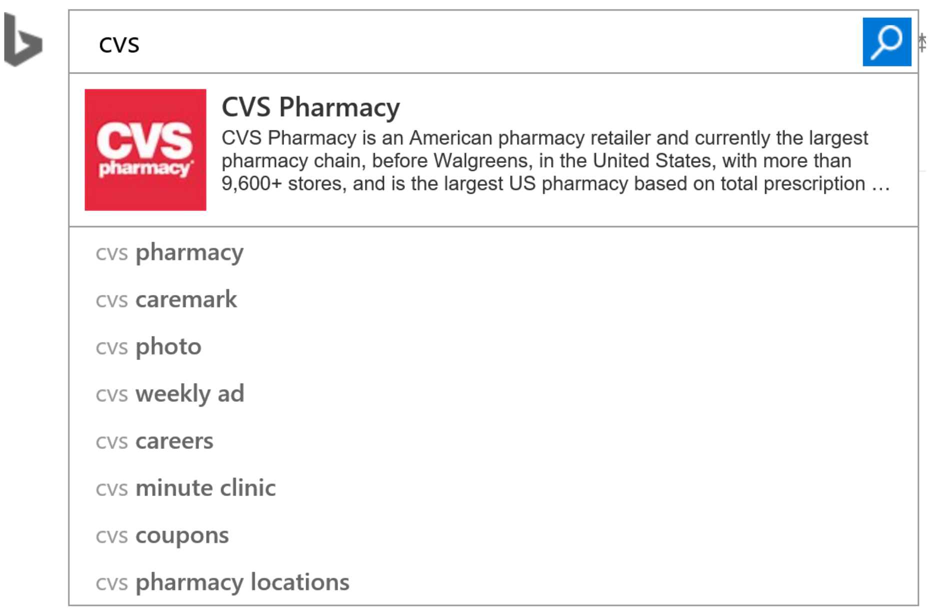
Query Understanding Framework

- Four steps

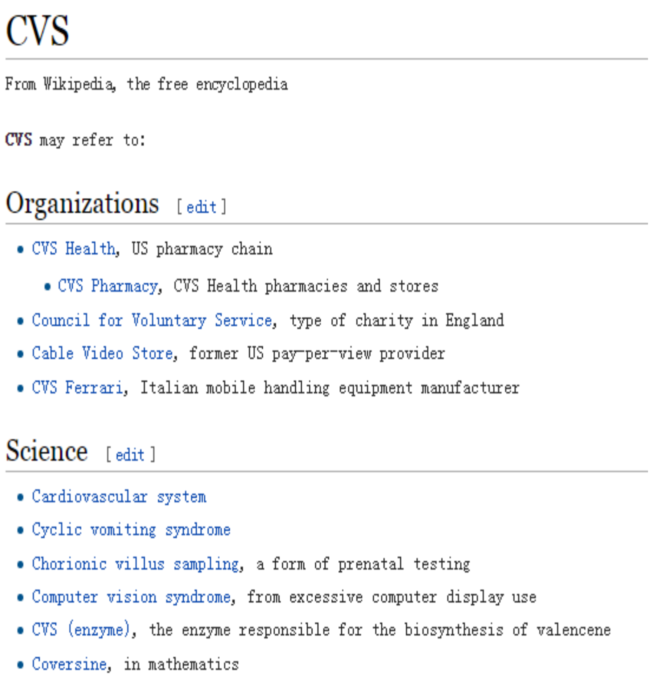


Step 1: Subtopic Candidate Mining

(1) Query Suggestions of Search Engines (Bing, Yahoo!, Google, and Baidu)



(2) Disambiguation items from KB (Wikipedia/Baidu Baike)


CVS

From Wikipedia, the free encyclopedia

CVS may refer to:

Organizations [edit]

- CVS Health, US pharmacy chain
  - CVS Pharmacy, CVS Health pharmacies and stores
- Council for Voluntary Service, type of charity in England
- Cable Video Store, former US pay-per-view provider
- CVS Ferrari, Italian mobile handling equipment manufacturer

Science [edit]

- Cardiovascular system
- Cyclic vomiting syndrome
- Chorionic villus sampling, a form of prenatal testing
- Computer vision syndrome, from excessive computer display use
- CVS (enzyme), the enzyme responsible for the biosynthesis of valencene
- Coversine, in mathematics

We didn't use query logs

Step 2: Candidate Clustering

- Original candidates gathered from different sources might be duplicated
  - cvs pharmacy, cvs health -> should be grouped into one subtopic
  - cvs version control software
- Clustering: group similar subtopic candidates into one cluster
  - Retrieve top 300 search results for each subtopic candidate
  - Generate a tf-idf vector based on snippets of the search results
  - Clustering algorithms: K-medoids, Quality Threshold (QT)

- A cluster -> a subtopic (the medoid)
  - {cvs pharmacy, cvs health} -> cvs pharmacy

Step 3: Subtopic Ranking

- Maximum Marginal Relevance (MMR) framework

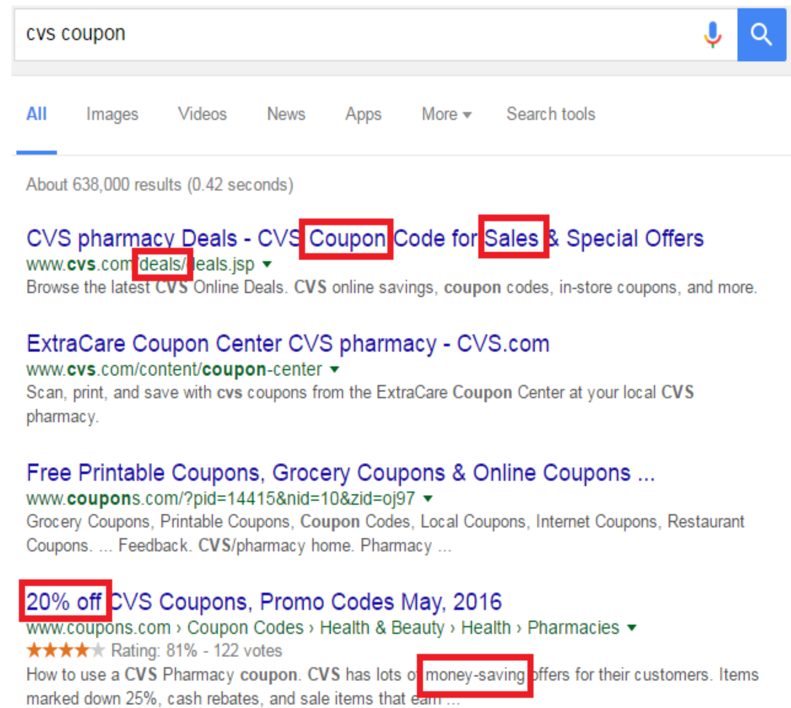
$$d_{i+1} = \arg\max\{\lambda Rel(d) + (1-\lambda)Div(d, D_i)\}$$

- Relevance - Rel(d): a linear combination of
  - Normalized size of the cluster : the subtopic is more important if the corresponding cluster contains more candidates (means the subtopic is more popular)
  - 1/subtopic length: a shorter subtopic is more general

- Novelty - Div(d, D): use maximun cosine similarity of tf-idf vectors between d and d_i in D(already ranked documents), to punish similar subtopics

Step 4: Vertical Classification (1)

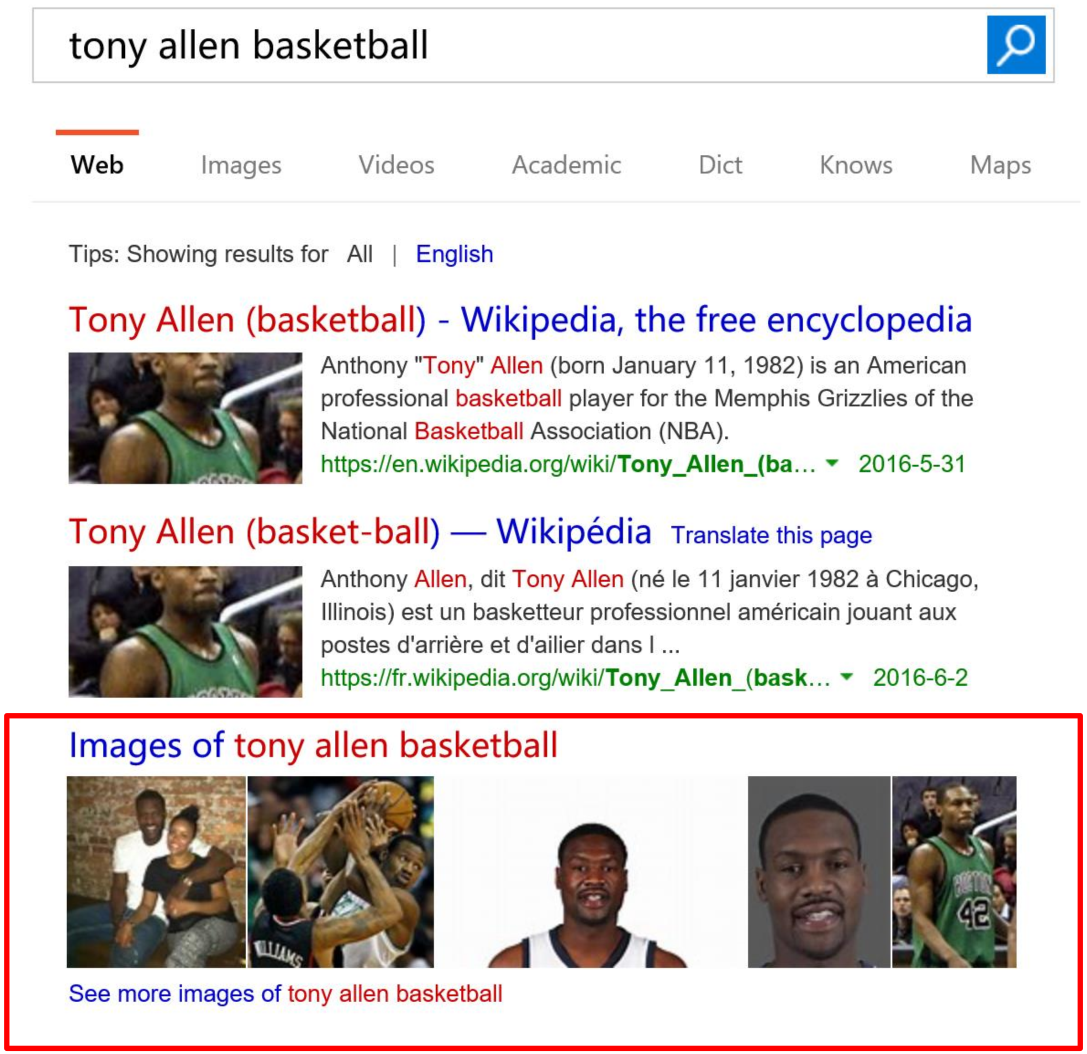
- A rule-based method
- Human created keyword-based rules
  - A subtopic belongs to a vertical if more than *50%* of top *100* results contain a specific keyword
  - e.g., Shopping: "coupon", "sale", "deal", "d% off"…..



Step 4: Vertical Classification (2)

- A supervised classification model
- Features:
  - Occurrences of human selected words within the top results
- Classification model: SVM (liblinear, L1-loss linear SVM)
- Training Data
  - Use IMine-1 topics and their query suggestions as training examples
  - Use Bing to generate training labels (except the Web vertical)

Tony allen basketball – image vertical



Using Bing to generate training labels:
For each query (subtopic), check whether Bing's SERP include answers/results from the specific vertical

Step 4: Vertical Classification (2)

- A separated model for each vertical
- Vertical is Web if no other verticals are returned

Runs and Results

Run Name	System Description	QU-score
rucir-Q-C/E-1Q	Suggestions + Wikipedia, k-medoids, trained classifier	**0.5757** **0.5613**
rucir-Q-C/E-2Q	Suggestions + Wikipedia, k-medoids, rule-based classifier	0.5495 0.5904
rucir-Q-C/E-3Q	Suggestions + Wikipedia, QT clustering, trained classifier	0.4489 0.4166
rucir-Q-C/E-4Q	Suggestions [redacted], k-medoids, trained classifier	0.5311 0.5583
rucir-Q-C/E-5Q	Suggestions + Ranking + Diversification	0.6849 0.6911

(1) Trained classifiers >Rule based (1>2)

(2) Query suggestions + Wikipedia > Query suggestions (1>4)

(3) Clustering algorithms: K-medoids > QT (1>3, 2>4)

Runs and Results

Run Name	System Description	QU-score
rucir-Q-C/E-1Q	Suggestions + Wikipedia, k-medoids, trained classifier	**0.5757** **0.5613**
rucir-Q-C/E-2Q	Suggestions + Wikipedia, k-medoids, rule-based classifier	0.5495 0.5904
rucir-Q-C/E-3Q	Suggestions + Wikipedia, QT clustering, trained classifier	0.4489 0.4166
rucir-Q-C/E-4Q	Suggestions [redacted], k-medoids, trained classifier	0.5311 0.5583
rucir-Q-C/E-5Q	Suggestions + Ranking + Diversification	0.6849 0.6911

!!!Direct ranking & diversification of original subtopic candidates > clustering (5>>1,2,3,4)

- Clustering: the selected subtopic failed to describe the entire cluster
- labelling problem: similar subtopics are identified as different subtopics by human (cvs pharmacy, cvs health, understand investigation)
- Other reasons (understand investigation)

NTCIR-12 IMine-2 Task

- **Goal: go toward whole-page diversification**
  - Diversify Web results together with other vertical answers
- **Subtask 1: Query Understanding**
  - Given a query, generate its subtopics, then classify their vertical intents
- **Subtask 2: Vertical Incorporating**
  - Diversify original search results together with virtual vertical documents

Diversification Model

- Directly use the PM2 Model
  - find the best subtopic based on current selected documents $$q_i = \frac{w_i}{2s_i + 1}$$
  - choose the next best document by the selected subtopic

$$d^* = \arg\max_{d \in D}[\lambda \cdot q^* \cdot rel(d, t^*) + (1 - \lambda) \cdot \sum_{t_i \neq t^*} q_i \cdot rel(d, t_i)]$$

$$s_i = s_i + \frac{rel(d^*, t_i)}{\sum_{rel(d,t_j)>0} rel(d^*, t_j)}$$

BM25 Model

Virtual Vertical Documents

- Each virtual vertical document is treated as a normal document
- Deal with rel(d, t) for virtual vertical documents
  - For each vertical, assume that its virtual document is "highly" relevant to the subtopics belongs to this vertical (rel(d*, t)=0.99) , and is irrelevant to other subtopics (rel(d*, t)=0)

Is it enough?

Search Result Diversification

- Our previous experiences
  - TREC 2009 & WSDM2011: multi-dimensional search result diversification (the best diversity run in TREC 2009)
  - CIKM 2015, SIGIR2016: Hierarchical search result diversification
  - JCST: search result diversification based on query facets
- What we have learned
  - A short string is not so effective for describing a subtopic/user intent
  - A single layer diversification might not be enough
  - More information is needed

Subtopic Expansion

- A simple method tried: use more keywords to help describe the user intent more completely and help find more diverse documents for a subtopic
- *Subtopic expansion based on* query suggestions
  - Retrieve Bing's query suggestions of the subtopic
    - *PlayStation 4 ->"PlayStation 4 new video game", "PlayStation 4 best price", and "PlayStation 4 controller"*
  - Combine all keywords as the expanded subtopic
    - *PlayStation 4 ->"PlayStation 4 video game best price controller"*
  - *Help return "diverse" documents for "playstation 4"*

Subtopics Used

- Original subtopics (and their verticals)
  - rucir-Q-C/E-1Q in the QU task
  - Official query suggestions provided by the organizers
- Expanded subtopics

Runs and Results

Table 3: Results of submitted runs for unclear queries in Vertical Incorporating subtask. The best result is in bold. Statistically significant differences among the submitted runs are marked with $\ast$, $\star$, $\circ$, $\triangle$, $\dagger$.

	Chinese Unclear Queries			English Unclear Queries		
Run Name	D♯-nDCG@10	D-nDCG@10	I-Recall	D♯-nDCG@10	D-nDCG@10	I-Recall
rucir-V-C/E-1M* [SExp+QU]	**0.7395**$^{\star\circ\triangle\dagger}$	**0.5342**$^{\circ\triangle\dagger}$	**0.9449**$^{\star\circ\triangle\dagger}$	**0.8249**$^{\star\circ\triangle\dagger}$	**0.6565**$^{\star\triangle\dagger}$	**0.9933**$^{\circ\triangle}$
rucir-V-C/E-2M* [SExp+Sug]	0.7079†	0.5268$^{\circ\triangle\dagger}$	0.8890	0.7807	0.5912	0.9701
rucir-V-C/E-3M$^{\circ}$ [noSExp+QU]	0.6884	0.4682	0.9086	0.7994	0.6534$^{\star\triangle\dagger}$	0.9454
rucir-V-C/E-4M$^{\triangle}$ [noSExp+Sug]	0.6801	0.4799	0.8802	0.7719	0.5847	0.9591
rucir-V-C/E-5M† [Baseline]	0.6593	0.4444	0.8742	0.7800	0.5723	0.9876

- Subtopic expansion is quite helpful
  - 1 > 3 (0.74 vs 0.69), 2 > 4
- Using QU subtopics is better than using pure suggestions
  - 1 > 2 (0.74 vs 0.71), 3 > 4

Summarization

- QU task
  - Trained classifiers are better than simple rules
  - Using wikipedia/baike data is a plus
- VI task
  - Go beyond the traditional single-layer subtopic based methods
  - Experimented with the subtopic expansion approach
- What we’ve learned
  - Subtopic clustering and selection might be risky

Thank you!

Zhicheng Dou
dou@ruc.edu.cn

Problems

- We retrieved lots of subtopic candidates, but only the selected subtopic is used and returned for diversification, other candidates are discarded – they should be used in the VI task