Overview of Contextual Suggestion Track

Heejun Kim
February 10, 2014

UNC
SCHOOL OF INFORMATION AND LIBRARY SCIENCE

INTRODUCTION

Goal & Some Changes from Previous year

- Examines search techniques for complicated information needs that are highly dependent on context and user interests
- Developes evaluation methodologies for such systems
- Changes
  - contexts will consist of only a location, no longer include a temporal component
  - Users from not only student, but also from crowdsourcing services
  - Suggestion attractiveness judgment with a 5-point scale

Goal & Some Changes

- Changes (continued..)
  - Submissions can be from either the open web or ClueWeb12
  - Profile, contexts, and returned suggestions will be formatted as JSON and CSV files rather than XML
  - A modified Time-Biased Gain (TBG) metric used in addition to P@5 and Mean Reciprocal Rank (MRR)

Data from Track

- A set of user profiles (Total: 562)
  - Reveals user's preference with respect to each sample suggestion
  - In 2013, the user preference indicates to a list of 50 example suggestions within Philadelphia, PA

- An example

id,title,description,url

1,Fresh on Bloor,"Our vegan menu boasts an array of exotic starters, multi-layered salads, filling wraps, high protein burgers and our signature Fresh bowls.",http://www.freshrestaurants.ca

id,attraction_id,description,website

1,1,1,0

1,2,3,4

Data from Track

- A set of geographical contexts
  - Corresponds to a particular location at with a city level granularity
  - 50 cities randomly selected in the US excluding Philadelphia, PA (the seed city)
- An example

id,city,state,lat,long

1,New York City,NY,40.71427,-74.00597

2,Chicago,IL,41.85003,-87.65005

Submitted Suggestions

- Consists of up to 50 ranked suggestions for each profile-context pair
- An example

groupid,runid,profile,context,rank,title,description,url,docId

group44,run44A,1,1,1,Deschutes Brewery Portland Public House,"Deschutes Brewery’s distinct Northwest brew pub in Portland’s

Pearl District has become a convivial gathering spot of beer and food lovers since it’s 2008 opening.",

http://www.deschutesbrewery.com,Data for retrieval

Either the open web or the ClueWeb12 dataset

Judgment

- Profile relevance
  - Top 5 ranked suggestions for each run for their profile among one or two randomly chosen profiles were rated
  - A 5 scale rating
  - UDInfoCS1 534 71 http://www.yelp.com/biz/cotton-monroe 2 3 31 13
- Geographical relevance
  - Whether the attraction was in the city or not
  - A 3 scale rating
  - 71 http://www.yelp.com/biz/waterfront-grill-monroe 2

UNC
SCHOOL OF INFORMATION AND LIBRARY SCIENCE

Judgment

- Baseline runs
  - BaselineA: top 50 attractions returned by the Google Places API from the open web
  - BaselineB: same strategy, but suggestions not in ClueWeb12 were filtered out
  - Only simply geographic context, not with user profile

EVALUATION METHODOLOGY

Measures

- P@5
  - How many of the top 5 ranked attraction are relevant
- Mean Reciprocal Rank (MRR)
  - 1/k, where k is the rank of the first relevant attraction found
- A modified Time-Biased Gain (TBG)

$$\sum_{k=1}^{5} D(T(k))A(k)(1-\Theta)^{\sum_{j=1}^{k-1} z(j)}$$

(from overview of the TREC 2013 Contextual Suggestion Track)

ALGORITHMIC SOLUTIONS

University of Lugano

- Objective
  - Concentrates on building user profiles very prudently to catch user preferences more precisely
  - In general, Natural Language Processing (NLP) technique and Machine Learning (ML) approach were used
- Four steps
  - Processing geo contexts
  - Inferring user term preferences
  - Building a personal ranking model
  - Ranking suggestions

University of Lugano

- Processing geographical contexts
  - Google Places API
  - Yandex Rich Content API (for generating description)
- Inferring User Term Preferences
  - Separate between positive and negative preferences
  - Extract only noun, adjectives, adverbs, and verbs by Natural Language Toolkit (NLTK)
  - Represent these words as binary vector
  - Expanding terms by introducing synonym dictionary (WordNet)
  - Calculate cosine distance between description and user profile

$$cos^{+}(\overrightarrow{D_i}, \overrightarrow{M_u^{+}}) = \frac{\overrightarrow{D_i} \cdot \overrightarrow{M_u^{+}}}{\|\overrightarrow{D_i}\| \cdot \|\overrightarrow{M_u^{+}}\|}$$

University of Lugano

- Building a personal ranking model
  - Weighted suggestion from the description and website rating

$$Weight(S) = \lambda R_{desc,s} + (1 - \lambda) R_{url,s}$$

  - Set thresholds to separate positive and negative samples in the training set
  - By default, λ = 0.5, but varies it depending on the kind of venue
  - Naïve Bayes classifier as a learning algorithm (Weka implementation)
- Ranking suggestions
  - Confidence score from the Naïve Bayes classifier

University of Lugano

- Result

Run	P@5	MRR	TBG
simpleScore	**0.4332**	**0.5871**	**1.8374**
complexScore	0.4152	0.5777	1.8226
median	0.2368	0.3415	0.8593

Table 2: Results for our two runs and median scores.

Run	P@5	MRR	TBG
simpleScore	21.97%	48.43%	13.90%
complexScore	21.97%	48.43%	15.70%

Table 3: Share of user/context pairs where particular run returned the best result over all entrants.

University of Delaware

- Objective
  - Evaluates an opinion-based method to model user profiles and rank candidate suggestions
  - Proposes template-based summarization method
- Two major contribution
  - For user modeling, estimate a positive user profile to model what the user likes, and vice versa
  - Leverage the above model by utilizing the on-line opinions posted by other real world users

University of Delaware

- System Framework

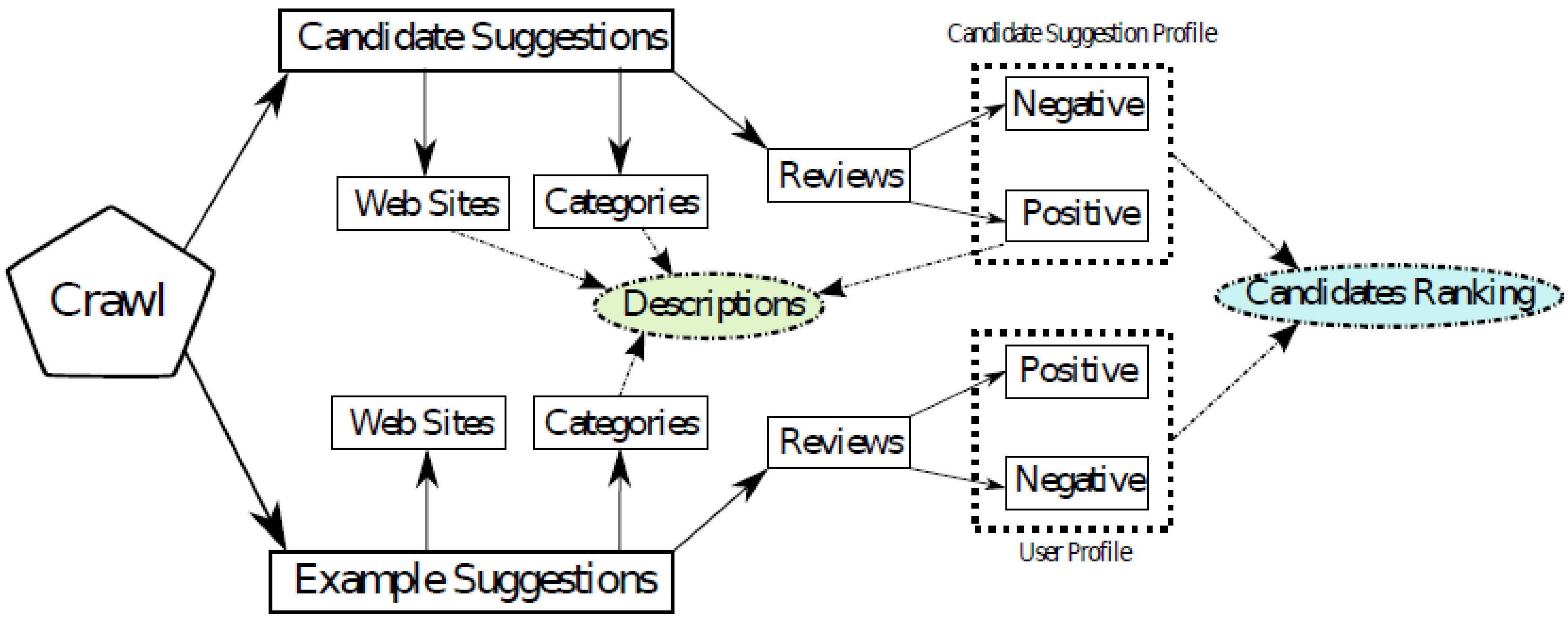


Fig. 1. System Framework

University of Delaware

- User profile Modeling

$$\mathcal{U}_{pos} = \bigcup_{es_i \in ES(U) \bigcap R_U(es_i)=POS} REP(O_{pos}(es_i))$$
$$\mathcal{U}_{neg} = \bigcup_{es_i \in ES(U) \bigcap R_U(es_i)=NEG} REP(O_{neg}(es_i))$$

$$\mathcal{CS}_{pos} = REP(O_{pos}(CS))$$
$$\mathcal{CS}_{neg} = REP(O_{neg}(CS)).$$

- Candidate suggestion ranking

$$S(U, CS) = \alpha \times SIM(\mathcal{U}_{pos}, \mathcal{CS}_{pos}) - \beta \times SIM(\mathcal{U}_{pos}, \mathcal{CS}_{neg}) - \gamma \times SIM(\mathcal{U}_{neg}, \mathcal{CS}_{pos}) + \eta \times SIM(\mathcal{U}_{neg}, \mathcal{CS}_{neg})$$

Democritus University of Thrace (DuTH)

- Context Processing

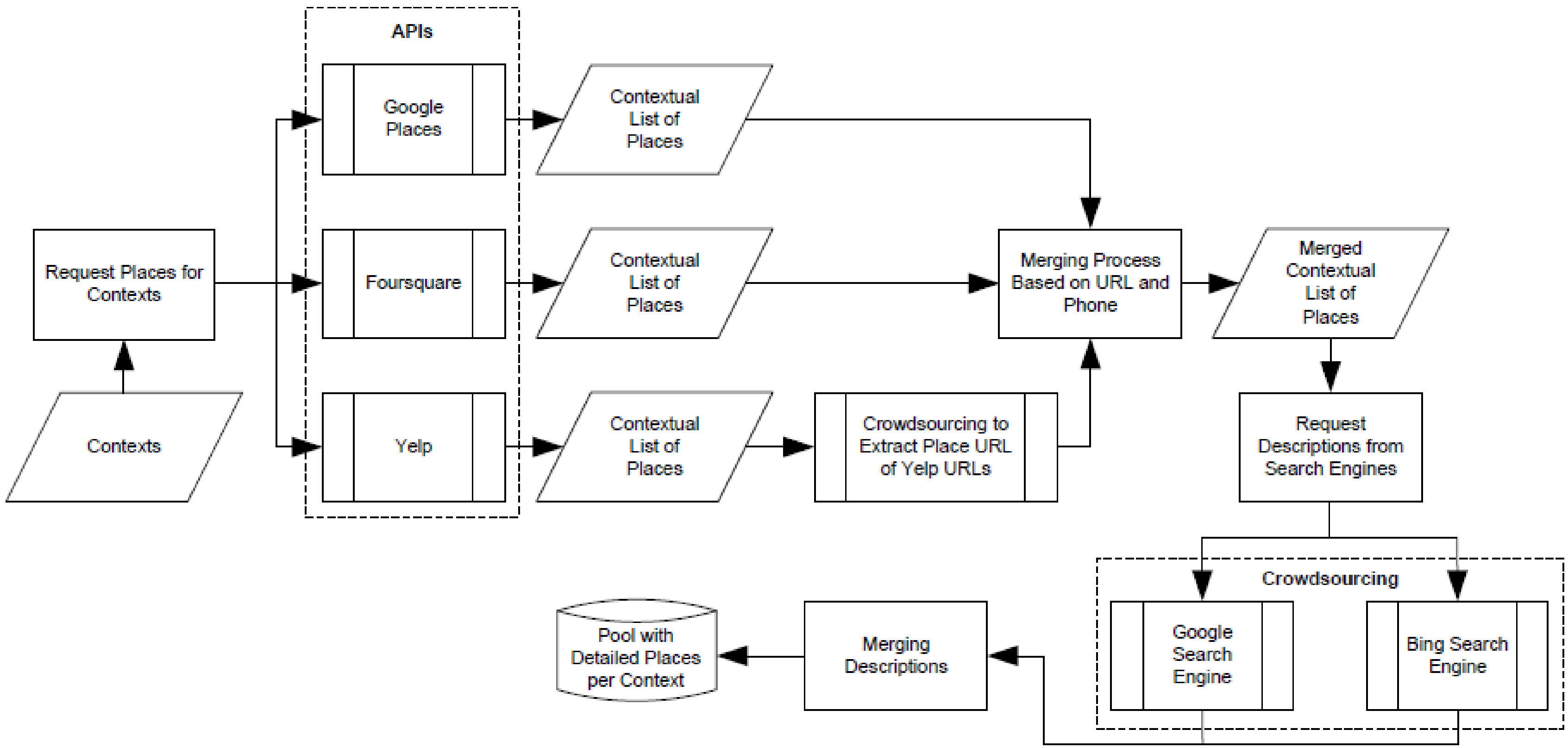
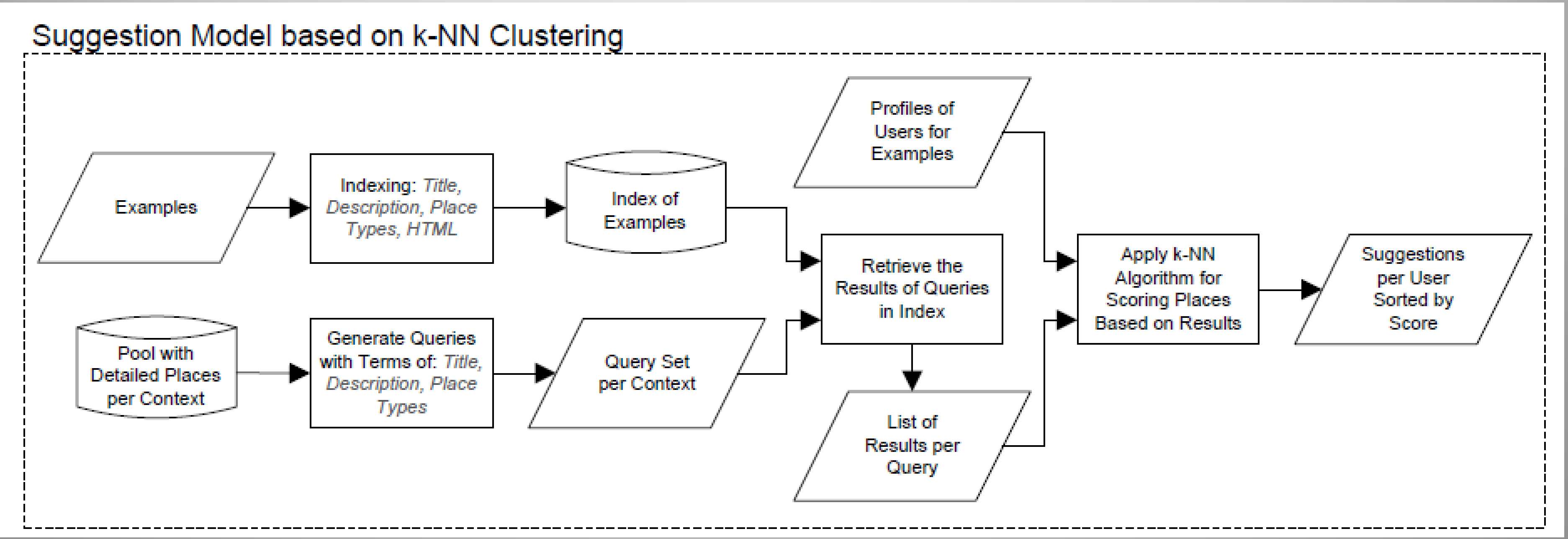


Figure 1: A system flowchart of the context processing.

Democritus University of Thrace (DuTH)

- Suggestion Processing



$$Q_{sc} = P_{sc} = \frac{\sum_{i=1}^{k} S_i \cdot avg(R_i^D, R_i^W)}{\sum_{i=1}^{k} S_i}$$

Democritus University of Thrace (DuTH)

- Suggestion Processing

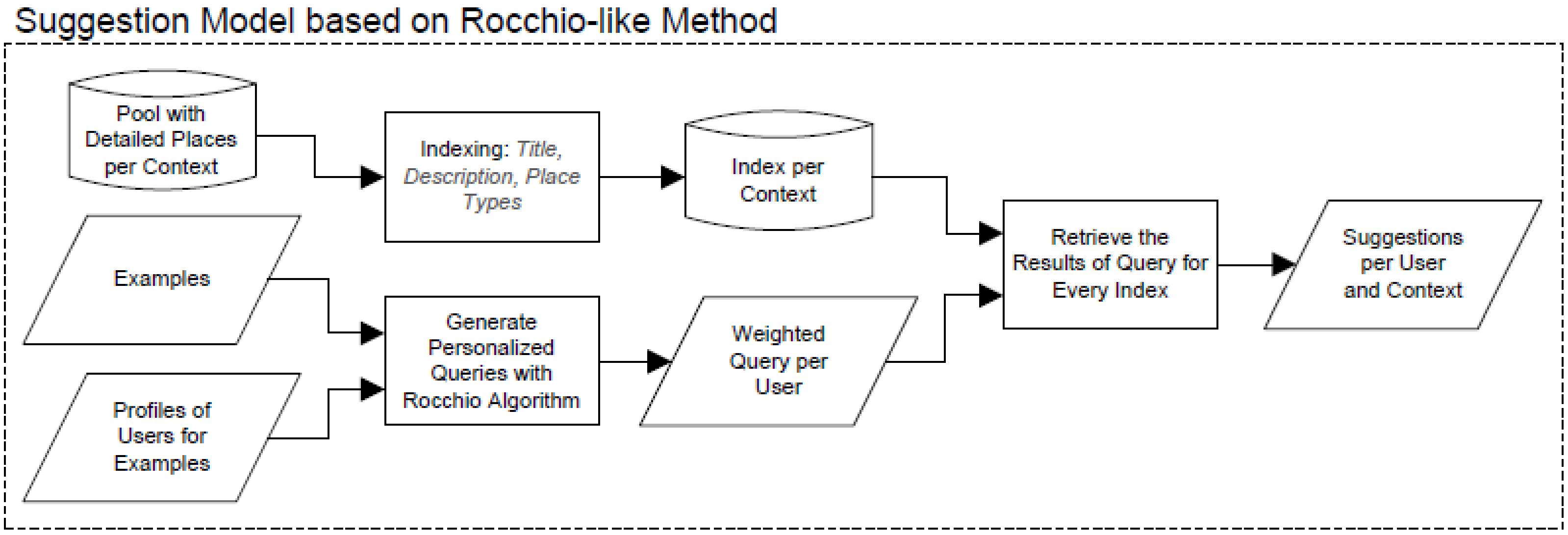


Figure 2: An overview of the proposed suggestion models.

$$Q_u = \sum_{j=0}^{4} \left((j-2) \frac{1}{|R_{j,u}|} \sum_{D \in R_{j,u}} D \right)$$

$$d_{i,j} = log(1 + f_{i,j})$$

Democritus University of Thrace (DUTH)

- Result

Table 2: Mean of results over all the profiles and contexts for P@5, MRR and TBG measures.

	P@5	MRR	TBG
Runs:			
DuTH_A	0.3283	0.4836	1.3109
DuTH_B	0.4090	0.5955	1.8508
Difference:			
DuTH_B vs _A	+24,58%	+23,14%	+41,19%

Table 3: Number of context-profile pairs with Median-or-better and Best scores per measure.

Runs	Median-or-better			Best		
	P@5	MRR	TBG	P@5	MRR	TBG
DuTH_A	189	175	151	25	86	22
DuTH_B	**209**	206	185	47	**114**	40
Total: 223 judged context-profile pairs						

COMPARISON OF THREE STUDIES

Processing Geographical Contexts

	Search API	Description
Lugano	Google Places API	Yandex Rich Content API
DUTH	Google Places API/ Foursquare/ Yelp	Crowdsourcing
University of Delaware	A crawler (not known) collects from the open web	Reviews of candidates from Yelp

Inferring User Preferences

	Search API	Tool used
Lugano	Extract words from profiles and build binary vector	NLTK
Lugano	Expanding words by synonym dictionary	WordNet
Lugano	Using Naïve Bayes classifier for learning algorithm	Weka
DuTH	k-NN clustering method	Indri v5.5 (indexing/search engine)
DuTH	Rocchio-like method	
University of Delaware	User profile modeling + Candidate suggestion modeling + similarity measurement	F2-EXP (retrieval model)

Result

	P@5	MRR	TBG
Lugano Simple Score	0.4332	0.5871	1.8374
Lugano Complex Score	0.4152	0.5777	1.8226
Duth k-NN clustering	0.3283	0.4836	1.3109
Duth Rocchio-like method	0.4090	0.5955	1.8508
University of Delaware 1	**0.5094**	**0.632**	**2.4474**
University of Delaware 2	0.4969	0.63	2.431
median	0.2368	0.3415	0.8593

BRAIN-STORMING TOPICS

Brain-Storming Topic

- In general, how was the context suggestions track?
- Inferring User Preferences
  - The work by U of Lugano became good thanks to word expansion, but simple cosine similarity is enough?
  - Duth: Any room to improve k-NN clustering method? Why do we have big different result on this group?
  - U of Delaware: Generalization vs Specialization? Generalization is still valid on this specialized domain
- Evaluation method
  - Do you find any room to improve?
  - Only crowdsourcing and student users + NIST assessors?