# Tailor knowledge graph for query understanding: linking intent topics by propagation

**Shi Zhao**
z.s@pku.edu.cn

**Yan Zhang**
zhy@cis.pku.edu.cn

Department of Machine Intelligence, Peking University, Beijing, China
Key Laboratory on Machine Perception, Ministry of Education, Beijing, China

## Abstract

Knowledge graphs are recently used for enriching query representations in an entity-aware way for the rich facts organized around entities in it. However, few of the methods pay attention to non-entity words and clicked websites in queries, which also help conveying user intent. In this paper, we tackle the problem of intent understanding with innovatively representing entity words, refiners and clicked urls as *intent topics* in a unified knowledge graph based framework, in a way to exploit and expand knowledge graph which we call 'tailor'. We collaboratively exploit global knowledge in knowledge graphs and local contexts in query log to initialize intent representation, then propagate the enriched features in a graph consisting of intent topics using an unsupervised algorithm. The experiments prove intent topics with knowledge graph enriched features significantly enhance intent understanding.

## 1 Introduction

Query understanding is the process of generating a representation which characterizes a user's search intent (Croft et al., 2010), which is of vital importance for information retrieval. However, users are remarkably laconic in describing their information needs due to anomalous state of knowledge (Belkin et al., 1982), resulting in vague and underspecified queries, which makes it especially difficult to understand and locate what they intended for in mountains of web data. The problem is often significantly compounded that people convey their intent rather in a series of behaviors called a search session than a single query, leaving a wealth of clues including query reformulations, page visits, dwell times, etc. What's more, as entities are taking center stage (Yin and Shah, 2010), string-level or phrase-level modeling of intent soon hits the bottleneck, calling for an entity-aware perspective.

Knowledge repositories, better known as knowledge graphs, such as Wikipedia, DBpedia and Freebase, have been recently utilized for enhancing query understanding for the large amounts of world knowledge they've harvested about entities and facts. A widely accepted way to use knowledge graph is tying queries with it by annotating entities in them, also known as entity linking.

However, information need is conveyed through more than entities. Quite a few non-entity words, aka refiners or modifiers, as well as many urls are barely included in knowledge graph, while they play an irreplaceable role in intent understanding. For example, a user may query *toyota*, *volvo* or just enter *car soup, cars for sale* and click *www.carsoup.com*, which should be encoded in a form that we could perceive their closeness in intent. That's why at-a-glance info cards about merely recognized entity in the query are far from enough and previous methods disregarding refiners and urls are too limited to cover queries in majority.

We move one step further to tailor knowledge graph for representing more than entity words. We collect refiners and clicked urls along with entity words and model intents they represent using knowledge graph based features. We use Freebase[1], one of the largest available knowledge graph, in our work and our method can be easily generalized to other knowledge repositories.

We put up an idea of *intent topic* which can be query words or urls, whether mean an entity or not, representing an atomic information need. We identify them with *intent features* by exploiting global knowledge in Freebase and local con-

---

[1]http://www.freebase.com

texts in query sessions. Notice the new concept here is distinguished from *query intent* or *query facet* in previous literature for it is in a holistic view, not specifically meaning subtopics around a certain query.

Our intuitive observations as follows inspire us to represent intent features with topics and domains in knowledge graph and propagate the enriched features in the intent topic graph.

1) Query words and urls within the same session tend to indicate the same query intent.
2) Intent topics sharing similar query intent often relate to similar topics in knowledge graph.
3) Knowledge graph domains sketch the query intent briefly.

Observation 1 indicates domain coherency within sessions is a good starting point to generate intent features, along with Observation 2 and 3 lay the basis of proximity that the propagation rely on.

To the best of our knowledge, we're the first to represent intent behind entity words, refiners and urls in a unified knowledge graph based framework, in a way to exploit and expand knowledge graph which we call 'tailor'.

Our contributions include:

- An innovative and unified framework to represent intent topics, whether they can directly link to an entity in knowledge graph or not.
- A novel algorithm to generate a specified intent topic graph, which enables learning intent features in an unsupervised propagation method.
- With intent topic graph we can better understand user intent conducting session-based contextualization and potentially find highly-related intent topic.

The rest of the paper is organized as follows. Section 2 tells our methods to map queries to Freebase and initialize intent features. Section 3 is about how we model intent topics in a unified graph and the propagation framework to learn intent features. We provide experiments and analysis in Section 4. Related work and conclusions are presented at the end of the paper.

## 2 Labeling intent topic nodes with Freebase-enriched features

In Freebase, facts around a certain topic and multi-faceted intents they reflect is more like a global domain distribution, what facet do users exactly intend for is difficult to locate until in a specified context, namely a query session.

We take a line in query log as a *query*, exhibiting an interaction with the search engine, including query words and page clicks. And a sequence of queries with a certain time interval constitute a *session*, completely conveying an information need.

In existing knowledge graph, only a small part of urls are contained in views of web pages beyond number online. Even for query words, we can merely get access to some of them, which we call *entity words* and the rest *refiners*. To avoid misunderstanding, the url intent topics in the following will specially refer to the clicks without directly matched concepts in knowledge graph, otherwise they'll be taken as entity intent topic.

In this section, we propose a framework of knowledge graph enriched representation of intent topics, the following propagation in Section3 bases on it.

### 2.1 Freebase as a knowledge graph

Freebase has over 39 million concepts, aka topics, about real-world entities like people, places and things stored as *nodes* in a graph. They're linked to each other with annotated egdes named as *property*. These *edges* actually represent *facts*. There are over a billion such facts or relations that make up the graph and they're all available for free. Properties are grouped into types, types are grouped into domains, which gives a broad view of knowledge in addtion to specific topics.

We can tap into Freebase through dump data or API[2]. In our work, we retrieve related Freebase topics with relevance scores for entity words via Freebase search API, which is based on combination of topic's inbound and outbound link counts in Freebase and Wikipedia as well as a popularity score computed by Google, and all the facts about a given topic through Freebase topic API. We use $\mathbb{T} = \{t_1, t_2, ...t_n\}$, $\mathbb{D} = \{d_1, d_2, ...d_n\}$ to denote all Freebase topics and domains used in our work.

### 2.2 Enriching entities and queries with Freebase

We represent a query's candidate intent topics by three sets, $E_q$, $R_q$, $C_q$, where $E_q$ includes entity words and clicks which have equivalents in Freebase, $R_q$ the refiner words and $C_q$ the rest clicks.

---

Global knowledge in Freebase can directly enrich each $e$ in $E_q$ with Freebase topics represented in vector $\mathbf{t}^e$, for each candidate topic there's a Freebase domain distribution vector $\mathbf{d}^t$. As for the rest in $R_q$ and $C_q$, they can learn features in later propagation process.

For any topic $t_i$ in $\mathbf{t}^e$, the relevance of entity words $e$ and knowledge graph topic $t_i$ is estimated as follows:

$$\mathbf{t}_i^e = \frac{RelevanceScore(e, t_i)}{\max_{t_j \in \mathbb{T}} RelevanceScore(e, t_j)} \quad (1)$$

And the domain vector $\mathbf{d}^{t_i}$ for $t_i$ is:

$$\mathbf{d}_j^{t_i} = \frac{pr(d_j|t_i)}{\sum_{d_k \in \mathbb{D}} pr(d_k|t_i)} \quad (2)$$

$$pr(d_j|t_i) = \frac{\# \ of \ links \ of \ t_i \ in \ domain \ d_j}{\# \ of \ all \ links \ in \ domain \ d_j} \quad (3)$$

Then we'll get a knowledge graph enriched intent description of the query by combining that of $e, r, c$.

$$\mathbf{t}_i^q = \sum_{e \in E_q} \mathbf{t}_i^e w_q(e) + \sum_{r \in R_q} \mathbf{t}_i^r w_q(r) + \sum_{c \in C_q} \mathbf{t}_i^c w_q(c) \quad (4)$$

$$w_q(e) = NCount_q(e)\varphi(e) \quad (5)$$

Here $\mathbf{t}^e \ \mathbf{t}^r \ \mathbf{t}^c$ correspond to the topic vector of each entity, refiner and click respectively. The weight indicates how dominant it is in conveying intent in the query. It is in proportion to the normalized count as well as each occurrence's quality denoted by $\varphi(e)$. Such as for entity words in Equation (5), the quality $\varphi(e)$ can be estimated with the help of entity linking methods, which describes the probability of $e$ as a candidate reference. That for clicks and refiners will be explained later.

The query's domain feature can be calculated as follows:

$$\mathbf{d}_i^q = \frac{\sum_{t_j \in \mathbb{T}} \mathbf{d}_i^{t_j} \mathbf{t}_j^q}{\sum_{d_k \in \mathbb{D}} \sum_{t_j \in \mathbb{T}} \mathbf{d}_k^{t_j} \mathbf{t}_j^q} \quad (6)$$

It describes the probability of query $q$ in domain $d_i$, in which $\mathbf{t}_j^q$ can be calculated by Equation (4) and $\mathbf{d}_i^{t_j}$ via facts around topic $t_j$ by Equation (2).

## 2.3 Contextualized intent depiction of sessions

The aforesaid enriched features we get about queries rely heavily on global knowledge in Freebase, reflecting prior distribution in the feature space. In this part, we derive a contextualized description of session intent in a local view by aggregating all the global knowledge we get about the session's queries. The ambiguity of a single query can be alleviated by looking at the dominant domain within the session.

The intent features $\mathbf{t}^s$ and $\mathbf{d}^s$ of session $s$ can be represented by computations on its query set $Q_s = \{q_1, q_2, ...q_n\}$ with time-order decay.

$$\mathbf{t}_i^s = \frac{\sum_{q \in Q_s} \mathbf{t}_i^q \alpha^{rank(q)}}{\sum_{t_j \in \mathbb{T}} \sum_{q \in Q_s} \mathbf{t}_j^q \alpha^{rank(q)}} \quad (7)$$

where we put an exponential decay controlled by decay factor $\alpha$. We get domain feature the same way as Equation (6).

We'll put up an unsupervised method of learning knowledge graph based intent representation of refiners and clicks in the following part.

## 3 Propagating intent features in the intent topic graph

In this section, our idea is to characterize entities, refiners and urls uniformly as intent topics, tailoring knowledge graph to intent topic graph so as to enrich representations by propagation.

### 3.1 Modeling intent topic graph

As in last section, with $\mathbf{d}^s$ featuring the context, candidate intent topics in sessions can make intent topic nodes now. We use the concept *intent topic* to stress words with local contexts tell a specified information need, thus making a node. Taking entity word *fl* as an example, it can be recognized as the topic *Florida* in Freebase, while the intent behind it can hardly be mapped to a single intent topic, such as *travel* domain in *hollywood fl*, *education* domain in *community college in florida*, and *florida department of health* actually convey intent in *government* domain.

So each intent topic node is identified with its name string and Freebase-enriched intent features $\mathbf{t}$ and $\mathbf{d}$. They're directly linked by co-occurring in the same line in the query log and implicitly related via intent features similarities, so that constitute a large graph $G = < V, E, W >$, where $\forall w \in W$ denotes an explicit edge weight and $\forall v \in V$ an intent topic. With intent topics and their relations modeled in a graph, we can better understand the query space so as to find the intended query faster. We realize it by aggregating massive sessions.

The implicit intent similarity $ISim$ of any node pair $n$ and $v$ can be encoded as follows.

$$ISim_{n,v} = \beta SSim_{n,v} + \gamma DSim_{n,v} + \eta TSim_{n,v}$$
(8)

where $SSim$ denotes the names' string similarity, $DSim$ the similarity of their domain feature and $TSim$ the topic vector similarity, with $\beta$, $\gamma$ and $\eta$ controlling the weight. The parameters may vary due to different scenarios. We just provide a framework of modeling nodes' intent features, which actually mirror their proximity in query intent.

To put it in more details, we use jaccard similarity for name shinglings and cosine similarity for domain and topic vector. As query log induced intent topic graph is of considerable large size, the pair-wise similarity is computationally prohibitive, hence we use Local Sensitive Hash (Indyk and Motwani, 1998) for each similarity metric so as to compute $ISim$ just in candidate set. We use random hyperplane based hash family proposed in (Charikar, 2002) and set the hash code dimension and hash table numbers empirically to ensure the number of nodes falling into each bucket is relatively stable.

### 3.2 Merging nodes

Although our idea of specifying intent topics by context better models the multi-facets of queries, it obviously also brings a sparse issue. For example, in one session user query *beep lyrics* and click *www.lyricsandsongs.com*, *lyrics* is tagged with the song *beep* and the musician *Pussycat Dolls*, in another scenario *lyrics* occurs with the song *what you know* and url *www.dapslyrics.com*, intents behind these two nodes are so similar that they should come into one, otherwise connections between the two intent-coherent urls may be lost.

To avoid that, we conduct a merge process to integrate nodes with exactly the same names and contexts into one, combing linked nodes and intent features together.

For a set of nearly duplicate nodes $\omega$ the calculation of new node's features can be written as:

$$\hat{\mathbf{t}} = \frac{\sum_{u \in \omega} \mathbf{t}^u}{|\omega|}$$
(9)

$$\hat{\mathbf{d}} = \frac{\sum_{u \in \omega} \mathbf{d}^u}{|\omega|}$$
(10)

In other words, we gather candidate nodes retrieved by LSH and then calculate $ISim$ for them

with $\eta$ setting to 0. Only node pairs with $ISim$ higher than a merge threshold $\theta$ can be seen as duplicates. The merge process is summarized in Algorithm 1.

---

**Algorithm 1:** Merging similar nodes

**Input**: $G = <V, E, W>, \beta, \gamma, \eta, \theta$
**Output**: $\hat{G} = <\hat{V}, \hat{E}, \hat{W}>$
**begin**
    Initialize $\Omega \leftarrow \varnothing$
    **for** $v \in V$ **do**
        Find dupset $\omega_v$ with $ISim_{\beta,\gamma,\eta}$
        **if** $\exists u \in V, \omega_u \in \Omega$ *and* $\omega_v \cap \omega_u \neq \varnothing$
        **then**
            $\omega_v \leftarrow \omega_v \cup \omega_u$
            Remove $\omega_u$ from $\Omega$
        Add $\omega_v$ to $\Omega$
    **for** $\omega \in \Omega$ **do**
        Merge nodes in $\omega$ into new node $\hat{v}$
        Update $G$ with replacing nodes in $\omega$
        with $\hat{v}$

---

### 3.3 Label propagation

We utilize knowledge graph induced intent features instead of manually labels as constraints to conduct label propagation(Zhu and Ghahramani, 2002). The idea is that node labels are propagated to nearby nodes via weighted edges until convergence, as highly weighted edges indicate high probability of sharing labels.

Nodes in our work have soft labels, where each dimension of intent features denotes a label, such as a topic or domain of knowledge graph. As described in aforesaid observations, it is intuitively reasonable to propagate on the basis of explicit edges and implicit intent similarities. We illustrate the propagation with topic feature, that of domain feature is similar.

We use matrix $\mathbf{Y}^t \in \mathbb{R}^{|V|*|\mathbb{T}|}$ to denote the intent topic graph's initial topic feature labels, with element $\mathbf{Y}_{ik}^t$ indicating node $v_i$'s relevance to $t_k$, wherer $t_k \in \mathbb{T}$. $\mathbf{Y}^t$ is initialized based on the results of the feature enriching step in Section 2, with no manually-labelled instances needed in our model. As only part of nodes can directly map to Freebase topics, those are initialized as labelled nodes, then propagate $\mathbf{t}$ to their linked neighbors. The number of unlabelled data is written as $u$, while that of labelled data $l$ and the total number of nodes $N$.

The transition matrix $T$ indicates the impact of nodes on each other. Note that here the $w_{ij}$ can be replaced by other similarity measures such as $ISim$ in Section 3.2.

$$T_{ij} = \frac{w_{ij}}{\sum_{k=1}^{N} w_{kj}} \quad (11)$$

Let $D$ denote an $N \times N$ diagonal matrix with $d_{ii} = \sum_j T_{ij}$. Then we can get a normalized version of transition matrix $P = D^{-1}T$.

The normalized transition matrix can be split into 4 sub-matrices.

$$P = \begin{bmatrix} P_{ll} & P_{lu} \\ P_{ul} & P_{uu} \end{bmatrix} \quad (12)$$

At each step, we propagate and clamp the labelled data and repeat until $Y$ converges, the propagation step can be written as:

$$\hat{Y}_u = P_{uu}Y_u + P_{ul}Y_l \quad (13)$$

As is shown in (Zhu and Ghahramani, 2002; Zhu et al., 2003) the solution to the propagation converges to:

$$\hat{Y}_u = (I - P_{uu})^{-1} P_{ul} Y_l \quad (14)$$

### 3.4 The propagation framework for intent features

We carry the propagation in an iterative process illustrated in Algorithm 2.

---

**Algorithm 2:** Intent feature propagation

**Input**: G, $Y_l^t$, $Y_l^d$
**Output**: $\hat{G}$, $\hat{Y}_u^t$, $\hat{Y}_u^d$
Initialize $Y_l^t$ $Y^d$ with results of Section2
**repeat**
    Merge similar nodes according to Algorithm 1
    Compute matrix $P$
    **repeat**
        $\hat{Y}_u^t = P_{uu}Y_u^t + P_{ul}Y_l^t$
    **until** *Convergence*;
    Recompute $\hat{P}$ with $\hat{Y}^t$
    **repeat**
        $\hat{Y}_u^d = \hat{P}_{uu}Y_u^d + \hat{P}_{ul}Y_l^d$
    **until** *Convergence*;
**until** *no dups*;

---

Since intent features include both domain vector and topic vector, we propagate them in an alternating way. At first we label nodes as described in Section 2, though missing refiners' and some urls' intent features, they are just used for initialization. Then we propagate Freebase topic features based on explicit edge weights, so that more nodes in intent topic graph have topic features now. Then fetching the learned topic features, we reinput it into domain feature propagation, which means we recalculate the transition matrix combining the implicit learned $TSim$ into edge weight, then propagate domain vector of labelled nodes through the graph. At each iteration, we first update $Y^t$, then input it to update $Y^d$, therefore merge near duplicate intent topics to update the whole graph.

## 4 Experiments

### 4.1 Data preparation

#### 4.1.1 Search logs

We use AOL search log data for experiments. It includes 20 million web queries collected covering 500K users over three months in 2006.

Table 1: The query set

| # of sessions | 35140 |
|---|---|
| # of queries | 271127 |
| # of users | 21378 |
| # of urls | 63019 |

We preprocess the query log by keeping urls occurring more than 3 times and queries with 2 to 40 characters, then extract sessions considering 25 minutes duration. While user session segmentation can be improved with more sophisticated algorithms, this simple low-cost heuristic performs adequately for our purposes. We then move on to map queries to Freebase and empirically filter sessions that are less entity-centric. We use an annotation tool especially for short text (Ferragina and Scaiella, 2012) called Tagme[3] to recognize entities and observe only 16% of all the queries are exactly an entity itself, which means most of queries do have refiner words to convey information need. To ensure the precision of recognized entities, we set a significant threshold and bottom line threshold, queries should have at least one recognized entity with a likelihood above significant level, and those below bottom line are ignored. They are 0.19 and 0.05 in our work, which may vary with entity recognition method. The normalized
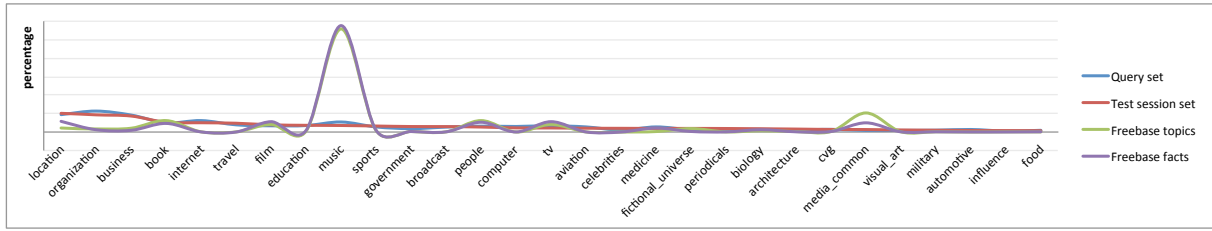
---

[3]http://tagme.di.unipi.it/

Figure 1: Unbalanced domain distributions in Freebase comparing against query set. Only domains with top proportions are shown.

Table 2: Examples of labelled intent topic nodes with learned feature

| Intent topic nodes | Original in Freebase | After propagation | Annotation |
|---|---|---|---|
| travel.yahoo.com | Yahoo! Travel (internet, 0.87), (projects, 0.13) | (location, 0.13), (travel, 0.11), (organization, 0.08), (business, 0.08) ... | Yahoo! Travel offers travel guides, booking and reservation services. |
| map quest www.mapquest.com | MapQuest (organization, 0.6), (book, 0.4) | (location, 0.13), (organization, 0.09), (travel, 0.09), (automotive, 0.06)... | MapQuest is an American free online web mapping service. |

likelihood is used as $w_q(e)$. Then we drop sessions where tagged entity words weight less than refiners as well as the ones with too many entity words spotted indicating disperse intents. For each recognized entity, only Freebase topics with relevance over 0.3 are kept. The query set we finally get is shown in Table 1.

### 4.1.2 Freebase

To enrich query representations, we collect a subset of Freebase including more than 7 millions facts and 4 millions topics in total which also contain 150 thousand topical equivalent websites, though less than 3% urls in query set are covered.

The facts and entities in Freebase is rather unbalanced across domains especially against that of recognized entities in query set as shown in Figure 1. Thus the original global knowledge we use about domain distribution may cause bias, which makes *tailoring* necessary for intent understanding.

For both generality and precision, we keep most of Freebase domains except several extreme incomplete ones, instead of retaining a small number of *representative* domains like many researchers do (Li et al., 2013; Yu et al., 2014; Lin et al., 2012). But generality comes at a price that some domains are confusing and mixed used which we then choose to merge, like celebrities and people,

periodicals and books, tv and broadcast, etc. We finally keep 50 of all 76 domains.

### 4.2 Intent topic graph

#### 4.2.1 Building the graph

We leverage both Freebase and search sessions to enrich intent topics. We set $\alpha$ to 0.9 in calculation of session's intent features. After labeling the session log, we roughly make a graph with 335206 intent topic nodes, 119364 of them have been labelled with Freebase topic feature, others only have domain feature. Then we conduct a merge process with $\beta$ set to 0.7, $\gamma$ to 0.3 and $\theta$ to 0.75 in order to merge nodes with duplicate names and similar contexts. We find 46659 duplicate sets covering 140768 nodes. Then we ignore nodes with few links and rare names to reduce sparsity. Finally we've got a graph of 209351 intent topics to initialize the propagation, including 78932 labelled nodes. The merge and propagation progress get converged in less than 4 rounds.

We'll further evaluate the graph with case study and a session intent understanding task.

#### 4.2.2 Case Study

We demonstrate intent features are good interpretations for query intent, whether they're labelled in Section 2 or learned by propagation in Section 3.

We can see in Table 2 that as nodes' original

Table 3: Examples of unlabelled intent topic nodes with learned feature

| Intent topic node | intent features | Annotation | Similarity nodes |
|---|---|---|---|
| www.bnm.com | (The Hertz Corporation, 0.25), (Southwest Florida International Airport, 0.17), (Punta Gorda Airport, 0.13), (Supercar, 0.09), (Sports car, 0.08)... (aviation, 0.23), (business, 0.21), (location, 0.14), (automotive, 0.11)... | Online booking of discount rentals at major airports, worldwide. | www.arac.com www.rentalcars.com www.hertz.com www.alamo.com rent a car cheap rental cars |
| www.mobtime.com | (Software, 0.18), (Mobile phone, 0.11), (100% Totally Free Ringtones, 0.10), (Motorola, 0.09), (Free Cell, 0.08), (Verizon Wireless, 0.04)... (computer, 0.23), (cvg, 0.21), (music, 0.19), (business, 0.11) ... | MobTime Cell Phone Manager is a PC software to manage or sync mobile phones. | cellphones.about.com cell software cell to pc reviews of cellphone wallpaper |

types in Freebase are not proper for describing intent, the intent features they get after propagation tend to be more explainable, such as the travel site often co-occurs with city names, tourist attractions, hotels and so on, thus indicating its intent in travel and location domain.

Table 3 shows examples which have no equivalents in Freebase. Although some of them may be accessible in other ontologies, we only take them as examples to show our propagation method makes it possible to depict intents behind urls and words in a knowledge graph based way while beyond the capacity of knowledge graph.

### 4.3 Session intent understanding task

#### 4.3.1 Experiment Setup

The evaluation of query understanding has long been a challenging task. To judge whether the concepts in query are successfully recognized seems too straightforward, and it can hardly be considered *understanding* the intent until the big idea about what *kind* of topics users emphasis is captured, which can be briefly sketched by distribution across Freebase domains. Also it is difficult to translate results of previous log analysis methods into knowledge graph domain information, thus hardly fit into our evaluation schema. We take popularity-based method as baseline.

We have few choices but to tag ground truth ourselves for intent understanding evaluation.

We randomly select 150 sessions as test set, the domain distribution of which agrees with the whole query set as shown in Figure 1. As mastering meanings of all Freebase domains is too challenging, we ask 5 accessors to describe each

session's intent broadly with a few natural language terms, then an expert familiar with Freebase schema translates the words into matched Freebase domains. Each test session is tagged by 2 accessors and 1 expert, we choose to use the tags of the cases in which the accessors reached agreement as the gold stantard. For example, if accessors tag session intent as pictures, then experts can translate it into Freebase visual art domain. Each session has 1∼4 tags and 1.6 tags in average. The tags cover 30 domains.

For each session, we derive the local intent domain vector $\mathbf{d}^s$ following the method in Section 2. Here we simply set quality function $\varphi(r)$ to a constant $\lambda^r$ for all refiners and $\varphi(c)$ to $\lambda^c$ for all clicks, we'll dive into more specialized weighting method in future work. $\lambda^r$ and $\lambda^c$ are parameters to control impact of different kinds of intent topics. Based on whether to exploit global intent features of non-entity words, we compare four variations against one baseline.

- Popularity-based (GP). We use domains' frequency in the query set as a baseline.
- Entity-based (E). We only use entity nodes' original intent features without propagation.
- Entity+Clicks (EC). Both intent features of entity words and clicks are used, controlled with $\lambda^c$.
- Entity+Refiners (ER). Intent features of entity words and refiner words are used, refiners' impact is controlled by $\lambda^r$.
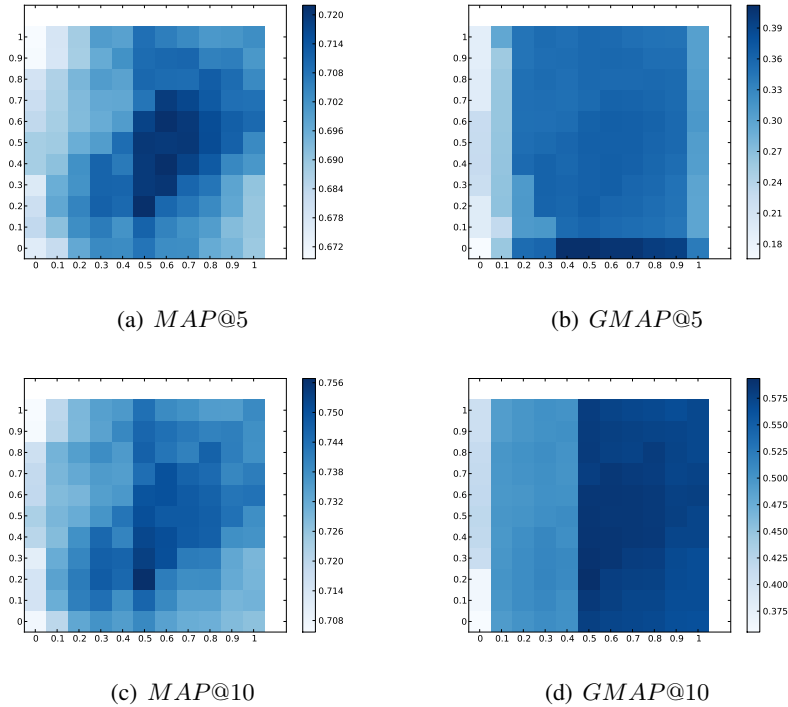- Entity+Clicks+Refiners (ECR). All intent topics are combined, controlled by $\lambda^c$, $\lambda^r$.

1076

(a) $MAP@5$      (b) $GMAP@5$

(c) $MAP@10$      (d) $GMAP@10$

Figure 2: The impact of $\lambda^r$ and $\lambda^c$ on ECR methods in four metrics, with vertical axis indicating $\lambda^r$, horizontal axis as $\lambda^c$. The first column on the left denotes ER method, while the bottom row the EC method.

### 4.3.2 Evaluation metrics

We use each approach to rank domains according to its derived weight, then compare it with golden standard set. It can be evaluated using Mean Average Precision (MAP), Geometric MAP and Precision@K. We use GMAP because it is more robust to outliers than the arithmetic mean.

For test set of size $N$, the MAP and GMAP can be calculated as follows:

$$MAP@k = \frac{1}{N} \sum_{i=1}^{N} AP_i@k \qquad (15)$$

$$GMAP@k = \sqrt[N]{\prod_{i=1}^{N} AP_i@k} \qquad (16)$$

### 4.3.3 Results and analysis

We first study impact of parameters $\lambda^r$ and $\lambda^c$, which is shown in Figure 2.

It roughly demonstrates different combinations of parameters' impact on ECR methods, performance is evaluated in four metrics, with deeper color indicating better result.

Best results comes with a $\lambda^c$ larger than $\lambda^r$ in all four subfigures. This trend seems more obvi-

ous in (d) where right part with larger $\lambda^c$ get better results. Also, deeper colors around diagonal line in (a) (c) indicate a more balanced combination of refiners and urls are more likely to enhance intent understanding. Thus we conclude clicks has a weak advantage over refiners in improving the result, while combining both with proper parameters can get the best result.

When comparing between MAP and GMAP, we can see while GMAP stays a high value when amplifying the impact of clicks, MAP changes with the variation of $\lambda^r$ for better or worse. As GMAP is a more robust metric, we can then infer that increasing weight of refiners could bring more outliers, implying refiners' intent features are more susceptible to noise.

Then we use ER with $\lambda^r = 0.5$ as $ER^{opt}$, EC with $\lambda^c = 0.5$ as $EC^{opt}$ and ECR with $\lambda^r = 0.2, \lambda^c = 0.5$ as $ECR^{opt}$.

Figure 3 clearly shows the superior performance of our model, especially at top positions. Table 4 shows the detailed comparisons between different methods. We can see our knowledge graph based intent representations perform well in session intent understanding. And refiners' and clicks' intent features which we learn by propagation con-
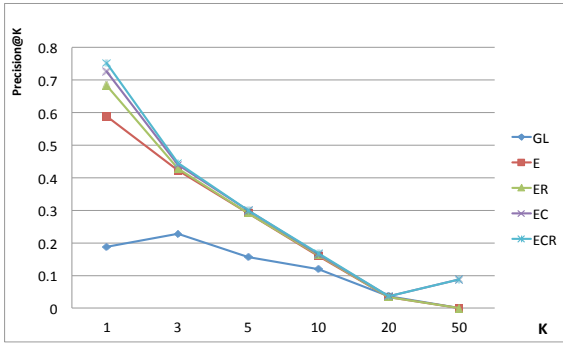
Figure 3: Precision@K results for different approaches, by varying number of k

Table 4: Comparisons among different methods

|  | K=5 | | K=10 | |
|---|---|---|---|---|
|  | MAP | GMAP | MAP | GMAP |
| $GP$ | 0.177 | 0.000 | 0.232 | 0.002 |
| $E$ | 0.676 | 0.166 | 0.707 | 0.355 |
| $EC^{opt}$ | 0.708 | 0.412 | 0.739 | 0.579 |
| $ER^{opt}$ | 0.688 | 0.227 | 0.723 | 0.421 |
| $ECR^{opt}$ | 0.722 | 0.412 | 0.756 | 0.594 |

tribute a lot to improve naive entity-based method, which do validate an complment effect of their learned intent features.

## 5 Related Work

### 5.1 Query intent understanding

Query intent or search intent has been studied intensively from various views.

A popular paradigm is to label several intents for each query, also called facets subgoals and subtopics in the literature, manully or by mining methods and then do classification (Hu et al., 2009; Li et al., 2008) based on that. Manually intent schemas range from 3 top level (Broder, 2002) to fine-grained subcatogories (Rose and Levinson, 2004) and taxonomy (Yin and Shah, 2010). Intent tasks in NTCIR-10 (Sakai et al., 2013) also provide subtopic pools made by accessors.

Another view of intent is more generic, mining or learning search intents without any kind of pre-defined intent category and clustering method is often used. Methods including (Sadikov et al., 2010; Yamamoto et al., 2012; Cheung and Li, 2012) cast intent as represented by a pattern or template consisting of a sequence of semantic concepts or lexical items. (Tan et al., 2012) encode intent in language models, aware of long-lasting interests. (Ren et al., 2014) uses an unsupervised

heterogeneous clustering. (Yin and Shah, 2010) capture generic intents around a certain named entities and model their relationships in a tree taxonomy and (Wang et al., 2009) mine broad latent modifiers of intent aspect , which are similar to our motivation, while we model more than intent phrases, but intent topics. We do not split queries into clusters or subtopics relevant to the original query to indicate a intent, but link them in an graph with intent feature similarity, weakly or strongly, in a holistical view.

On the other hand, previous research can be categorized by what kind of resources they rely on. Quite an amount of work leverage query logs (Jiang et al., 2013), including query reformulations (Radlinski et al., 2010), click-through data (Li et al., 2008). There are also works using sponsered data (Yamamoto et al., 2012) and interactive data (Ruotsalo et al., 2013). The new trend of integrating knowledge graph will be discussed next.

### 5.2 Knowledge graph on intent understanding

Instead of summarizing queries into concepts by clustering, recently there appears a tendency to use concpets from knowledge graph resources. Some researchers manage to build entity graph from queries (Bordino et al., 2013a) (Bordino et al., 2013b; Yu et al., 2014), some in a structure view, interpret quries into knowledge base fit template (Pound et al., 2012; Li et al., 2013). (Pantel et al., 2012) models latent intent to mine entity type distributions. (Ren et al., 2014) utilizes knowledge graph resources in a hetrogeneous view. (Lin et al., 2012) also pays attention to refiners, but restricted to limited domains, while our method is more general.

## 6 Conclusion

In this paper, we tailor knowledge graph to represent query intent behind entity words, refiners and clicked urls in a unified framework, taking them as intent topic nodes connected in a large graph. We manage to get a contextualized intent depiction exploiting global knowledge in Freebase, then propagate the feature to cover more intent topics. We show in experiments the knowledge graph enriched representation is reasonable and explainable, and the intents feature of refiners and clicks can better enhance intent understanding than methods simply relying on entities.

There are several directions for future work, including using both types and domains in Freebase schema, diving into refiners and looking for a proper weighting method, developing a query recommendation framework based on the intent topic graph and user interest modeling.

## Acknowledgments

## References

Nicholas J Belkin, Robert N Oddy, and Helen M Brooks. 1982. Ask for information retrieval: Part i. background and theory. *Journal of documentation*, 38(2):61–71.

Ilaria Bordino, Gianmarco De Francisci Morales, Ingmar Weber, and Francesco Bonchi. 2013a. From machu_picchu to rafting the urubamba river: anticipating information needs via the entity-query graph. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 275–284. ACM.

Ilaria Bordino, Yelena Mejova, and Mounia Lalmas. 2013b. Penguins in sweaters, or serendipitous entity search on user-generated content. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 109–118. ACM.

Andrei Broder. 2002. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, September.

Moses S Charikar. 2002. Similarity estimation techniques from rounding algorithms. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 380–388. ACM.

Jackie Chi Kit Cheung and Xiao Li. 2012. Sequence clustering and labeling for unsupervised query intent discovery. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 383–392. ACM.

W Bruce Croft, Michael Bendersky, Hang Li, and Gu Xu. 2010. Query representation and understanding workshop. In *SIGIR Forum*, volume 44, pages 48–53.

Paolo Ferragina and Ugo Scaiella. 2012. Fast and accurate annotation of short texts with wikipedia pages. *IEEE software*, 29(1).

Jian Hu, Gang Wang, Fred Lochovsky, Jian-tao Sun, and Zheng Chen. 2009. Understanding user's query intent with wikipedia. In *Proceedings of the 18th international conference on World wide web*, pages 471–480. ACM.

Piotr Indyk and Rajeev Motwani. 1998. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 604–613. ACM.

Daxin Jiang, Jian Pei, and Hang Li. 2013. Mining search and browse logs for web search: A survey. *ACM Trans. Intell. Syst. Technol.*, 4(4):57:1–57:37, October.

Xiao Li, Ye-Yi Wang, and Alex Acero. 2008. Learning query intent from regularized click graphs. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, pages 339–346, New York, NY, USA. ACM.

Yanen Li, Bo-June Paul Hsu, and ChengXiang Zhai. 2013. Unsupervised identification of synonymous query intent templates for attribute intents. In *Proceedings of the 22nd ACM international conference on Conference on information &#38; knowledge management*, CIKM '13, pages 2029–2038, New York, NY, USA. ACM.

Thomas Lin, Patrick Pantel, Michael Gamon, Anitha Kannan, and Ariel Fuxman. 2012. Active objects: Actions for entity-centric search. In *Proceedings of the 21st international conference on World Wide Web*, pages 589–598. ACM.

Patrick Pantel, Thomas Lin, and Michael Gamon. 2012. Mining entity types from query logs via user intent modeling. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 563–571. Association for Computational Linguistics.

Jeffrey Pound, Alexander K Hudek, Ihab F Ilyas, and Grant Weddell. 2012. Interpreting keyword queries over web knowledge bases. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 305–314. ACM.

Filip Radlinski, Martin Szummer, and Nick Craswell. 2010. Inferring query intent from reformulations and clicks. In *Proceedings of the 19th international conference on World wide web*, pages 1171–1172. ACM.

Xiang Ren, Yujing Wang, Xiao Yu, Jun Yan, Zheng Chen, and Jiawei Han. 2014. Heterogeneous graph-based intent learning with queries, web pages and wikipedia concepts. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 23–32, New York, NY, USA. ACM.

Daniel E. Rose and Danny Levinson. 2004. Understanding user goals in web search. In *Proceedings of the 13th International Conference on World*

*Wide Web*, WWW '04, pages 13–19, New York, NY, USA. ACM.

Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Głowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, et al. 2013. Directing exploratory search with interactive intent modeling. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1759–1764. ACM.

Eldar Sadikov, Jayant Madhavan, Lu Wang, and Alon Halevy. 2010. Clustering query refinements by user intent. In *Proceedings of the 19th international conference on World wide web*, pages 841–850. ACM.

Tetsuya Sakai, Zhicheng Dou, Takehiro Yamamoto, Yiqun Liu, Min Zhang, Ruihua Song, MP Kato, and M Iwata. 2013. Overview of the ntcir-10 intent-2 task. *Proceedings of NTCIR-10*, pages 94–123.

Bin Tan, Yuanhua Lv, and ChengXiang Zhai. 2012. Mining long-lasting exploratory user interests from search history. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1477–1481. ACM.

Xuanhui Wang, Deepayan Chakrabarti, and Kunal Punera. 2009. Mining broad latent query aspects from search sessions. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 867–876. ACM.

Takehiro Yamamoto, Tetsuya Sakai, Mayu Iwata, Chen Yu, Ji-Rong Wen, and Katsumi Tanaka. 2012. The wisdom of advertisers: mining subgoals via query clustering. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 505–514. ACM.

Xiaoxin Yin and Sarthak Shah. 2010. Building taxonomy of web search intents for name entity queries. In *Proceedings of the 19th international conference on World wide web*, pages 1001–1010. ACM.

Xiao Yu, Hao Ma, Bo-June (Paul) Hsu, and Jiawei Han. 2014. On building entity recommender systems using user click log and freebase knowledge. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, pages 263–272, New York, NY, USA. ACM.

Xiaojin Zhu and Zoubin Ghahramani. 2002. Learning from labeled and unlabeled data with label propagation. Technical report, Technical Report CMU-CALD-02-107, Carnegie Mellon University.

Xiaojin Zhu, Zoubin Ghahramani, John Lafferty, et al. 2003. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, volume 3, pages 912–919.