

# GIANT: Scalable Creation of a Web-scale Ontology

Bang Liu<sup>1\*</sup>, Weidong Guo<sup>2\*</sup>, Di Niu<sup>1</sup>, Jinwen Luo<sup>2</sup>  
Chaoyue Wang<sup>2</sup>, Zhen Wen<sup>2</sup>, Yu Xu<sup>2</sup>  
<sup>1</sup>University of Alberta, Edmonton, AB, Canada  
<sup>2</sup>Platform and Content Group, Tencent, Shenzhen, China

## ABSTRACT

Understanding what online users may pay attention to is key to content recommendation and search services. These services will benefit from a highly structured and web-scale ontology of entities, concepts, events, topics and categories. While existing knowledge bases and taxonomies embody a large volume of entities and categories, we argue that they fail to discover properly grained concepts, events and topics in the language style of online population. Neither is a logically structured ontology maintained among these notions. In this paper, we present *GIANT*, a mechanism to construct a user-centered, web-scale, structured ontology, containing a large number of natural language phrases conforming to user attentions at various granularities, mined from a vast volume of web documents and search click graphs. Various types of edges are also constructed to maintain a hierarchy in the ontology. We present our graph-neural-network-based techniques used in *GIANT*, and evaluate the proposed methods as compared to a variety of baselines. *GIANT* has produced the Attention Ontology, which has been deployed in various Tencent applications involving over a billion users. Online A/B testing performed on Tencent QQ Browser shows that Attention Ontology can significantly improve click-through rates in news recommendation.

## CCS CONCEPTS

• Information systems → Data mining; Document representation; • Computing methodologies → Information extraction.

\*Equal contribution. Correspondence to: weidongguo@tencent.com.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org). *SIGMOD'20, June 14–19, 2020, Portland, OR, USA*  
© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6735-6/20/06...\$15.00

<https://doi.org/10.1145/3318464.3386145>

## KEYWORDS

Ontology Creation, Concept Mining, Event Mining, User Interest Modeling, Document Understanding

## ACM Reference Format:

Bang Liu<sup>1\*</sup>, Weidong Guo<sup>2\*</sup>, Di Niu<sup>1</sup>, Jinwen Luo<sup>2</sup> and Chaoyue Wang<sup>2</sup>, Zhen Wen<sup>2</sup>, Yu Xu<sup>2</sup>. 2020. GIANT: Scalable Creation of a Web-scale Ontology. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD'20)*, June 14–19, 2020, Portland, OR, USA. ACM, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3318464.3386145>

## 1 INTRODUCTION

In a fast-paced society, most people carefully choose what they pay attention to in their overstimulated daily lives. With today's information explosion, it has become increasingly challenging to increase the attention span of online users. A variety of recommendation services [1, 2, 6, 28, 30] have been designed and built to recommend relevant information to online users based on their search queries or viewing histories. Despite a variety of innovations and efforts that have been made, these systems still suffer from two long-standing problems, inaccurate recommendation and monotonous recommendation.

Inaccurate recommendation is mainly attributed to the fact that most content recommender, e.g., news recommender, are based on keyword matching. For example, if a user reads an article on "Theresa May's resignation speech", current news recommenders will try to further retrieve articles that contain the keyword "Theresa May", although the user is most likely not interested in the person "Theresa May", but instead is actually concerned with "Brexit negotiation" as a topic, to which the event "Theresa May's resignation speech" belongs. Therefore, keywords that appear in an article may not always be able to characterize a user's interest. Frequently, a higher and proper level of abstraction of verbatim keywords is helpful to recommendation, e.g., knowing that Honda Civic is an "economy car" or "fuel-efficient car" is more helpful than knowing it is a sedan.

Monotonous recommendation is the scenario where users are recommended with articles that always describe the same entities or events repeatedly. This phenomenon is rooted in the incapability of existing systems to extrapolate beyond

the verbatim meaning of a viewed article. Taking the above example again, instead of pushing similar news articles about “Theresa May’s resignation speech” to the same user, a way to break out from redundant monotonous recommendation is to recognize that “Theresa May’s resignation speech” is an event in the bigger topic “Brexit negotiations” and find other preceding or follow-up events in the same topic.

To overcome the above-mentioned deficiencies, what is required is an ontology of “user-centered” terms discovered at the web scale that can abstract keywords into concepts and events into topics in the vast open domain, such that user interests can be represented and characterized at different granularities, while maintaining a structured hierarchy among these terms to facilitate interest inference. However, constructing an ontology of user interests, including entities, concepts, events, topics and categories, from the open web is a very challenging task. Most existing taxonomy or knowledge bases, such as Probase [69], DBPedia [32], CN-DBPedia [70], YAGO [63], extract concepts and entities from Wikipedia and web pages based on Hearst patterns, e.g., by mining phrases around “such as”, “especially”, “including” etc. However, concepts extracted this way are clearly limited. Moreover, web pages like Wikipedia are written in an author’s perspective and are not the best at tracking user-centered Internet phrases like “top-5 restaurants for families”.

To mine events, most existing methods [3, 40, 43, 72] rely on the *ACE (Automatic Content Extraction) definition* [15, 20] and perform event extraction within specific domains via supervised learning by following predefined patterns of triggers and arguments. This is not applicable to vastly diverse types of events in the open domain. There are also works attempting to extract events from social networks such as Twitter [4, 12, 55, 65]. But they represent events by clusters of keywords or phrases, without identifying a clear hierarchy or ontology among phrases.

In this paper, we propose **GIANT**, a web-based, structured ontology construction mechanism that can automatically discover critical natural language phrases or terms, which we call user *attentions*, that characterize user interests at different granularities, by mining unstructured web documents and search query logs at the web scale. In the meantime, GIANT also aims to form a hierarchy of the discovered user attention phrases to facilitate inference and extrapolation. GIANT produces and maintains a web-scale ontology, named the *Attention Ontology (AO)*, which consists of around 2 million nodes of five types, including categories, concepts, entities, topics, and events, and is still growing. In addition to categories (e.g., cars, technology) and entities (e.g., iPhone XS, Honda Civic) that can be commonly found in existing taxonomies or knowledge graphs, Attention Ontology also contains abstractive terms at various semantic granularities,

including newly discovered *concepts*, *events* and *topics*, all in user language or popular online terms. By tagging documents with these abstractive terms that the online population would actually pay attention to, Attention Ontology proves to be effective in improving content recommendation in the open domain. For example, with the ontology, the system can extrapolate onto the concept “economy cars” if a user has browsed an article about “Honda Civic”, even if the exact wording of “economy cars” does not appear in that article. The system can also infer a user’s interest in all events related to the topic “Brexit Negotiation” if he or she viewed an article about “Theresa May’s resignation”.

GIANT constructs the user-centered Attention Ontology by mining the *click graph*, a large bipartite graph formed by search queries and their corresponding clicked documents. GIANT relies on the linkage from queries to the clicked documents to discover terms that may represent user attentions. For example, if a query “top-5 family restaurants in Vancouver” frequently leads to the clicks on certain restaurants, we can recognize “top-5 family restaurants in Vancouver” as a concept and the clicked restaurants are entities within this concept. Compared to existing knowledge graphs constructed from Wikipedia, which contain general factual knowledge, queries from users can reflect hot concepts, events or topics that are of user interests at the present time.

To automatically extract concepts, events or topics from the click graph, we propose GCTSP-Net (Graph Convolution-Traveling Salesman Problem Network), a multi-task model which can mine different types of phrases at scale. Our model is able to extract heterogeneous phrases in a unified manner. The extracted phrases become the nodes in the Attention Ontology, which can properly summarize true user interests and can also be utilized to characterize the main theme of queries and documents. Furthermore, to maintain a hierarchy within the ontology, we propose various machine learning and ad-hoc approaches to identify the edges between nodes in the Attention Ontology and tag each edge with one of the three types of relationships: *isA*, *involve*, and *correlate*. The constructed edges can benefit a variety of applications, e.g., concept-tagging for documents at an abstractive level, query conceptualization, event-series tracking, etc.

We have performed extensive evaluation of GIANT. For all the learning components in GIANT, we introduce efficient strategies to quickly build the training data necessary to perform phrase mining and relation identification, with minimum manual labelling efforts. For phrase mining, we combine a bootstrapping strategy based on pattern-phrase duality [7, 34] with text alignment based on query-title conformity. For relationship identification, we utilize the co-occurrence of different phrases in queries, documents, and consecutive queries to extract phrase pairs labeled with

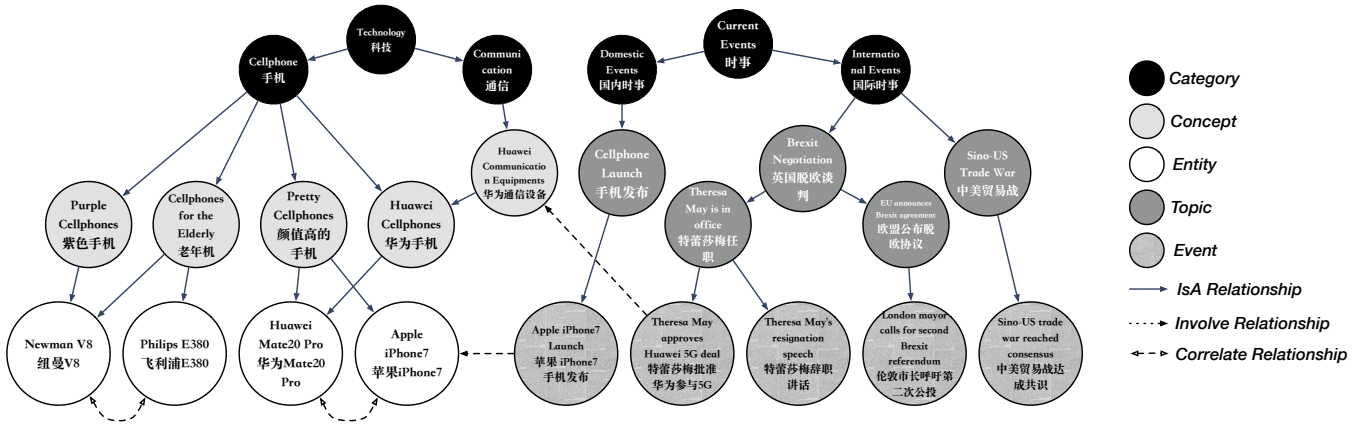


Figure 1: An example to illustrate our Attention Ontology (AO) for user-centered text understanding.

target relationships. These labeled examples automatically mined from the click graph are then used to train the proposed machine learning models in different sub-tasks.

The assessment of the constructed ontology shows that our system can extract a plethora of high-quality phrases, with a large amount of correctly identified relationships between these phrases. We compare the proposed GNN-based GCTSP-Net with a variety of baseline approaches to evaluate its performance and superiority on multiple tasks. The experimental results show that our approach can extract heterogeneous phrases more accurately from the click graph as compared to existing methods.

It is worth noting that GIANT has been deployed in multiple real-world applications, including Tencent QQ browser, Mobile QQ and WeChat, involving more than 1 billion active users around the globe, and currently serves as the core taxonomy construction system in these commercial applications to discover long-term and trending user attentions. We report the online A/B testing results of introducing Attention Ontology into Tencent QQ Browser mobile app, which is a news feed app that displays news articles as a user scrolls down on her phone. The results suggest that the generated Attention Ontology can significantly improve the click-through rate in news feeds recommendation.

## 2 THE ATTENTION ONTOLOGY

In the proposed Attention Ontology, each node is represented by a phrase or a collection of phrases of the same meaning mined from Internet. We use the term “attention” as a general name for *entities*, *concepts*, *events*, *topics*, and *categories*, which represent five types of information that can capture an online user’s attention at different semantic granularities. Such attention phrases can be utilized to conceptualize user interests and depict document coverages. For instance, if a user wants to buy a family vehicle for road trips, he/she may input such a query “vehicles choices for family road

trip”. From this query, we could extract the concept, “family road trip vehicles”, and tag it to matching entities such as “Honda Odyssey Minivan” or “Ford Edge SUV”. We could then recommend articles related to these Honda and Ford vehicles, even if they do not contain the exact wording of “family road trip”. In essence, the Attention Ontology enables us to achieve a user-centered understanding of web documents and queries, which improves the performance of search engines and recommender systems.

Figure 1 shows an illustration of the Attention Ontology, which is in the form of a Directed Acyclic Graph (DAG). Specifically, there are five types of nodes:

- **Category:** a category node defines a broad field that encapsulates many related topics or concepts. For example, technology, current events, entertainment, sports, finance and so forth. In our system, we pre-define a 3-level categories hierarchy, which consists of 1,206 different categories.
- **Concept:** a concept is a group of entities that share some common attributes [34, 69], such as superheroes, MOBA games, fuel-efficient cars and so on. In contrast with coarse-grained categories and fine-grained entities, concepts can help better characterize users’ interests at a suitable semantic granularity.
- **Entity:** an entity is a specific instance belonging to one or multiple concepts. For example, Iron Man is an entity belonging to the concepts “superheroes” and “Marvel superheroes”.
- **Event:** an event is a real-world incident that involves a group of specific persons, organizations, or entities. It is also tagged with a certain time/location of occurrence [35]. In our work, we represent an event with four types of attributes: *entities* (indicating who or what are involved in the event), *triggers* (indicating what kind/type of event it is), *time* (indicating when

the event takes place), and *location* (indicating where the event takes place).

- **Topic**: a topic represents a collection of events sharing some common attributes. For example, both “Samsung Galaxy Note 7 Explosion in China” and “iPhone 6 Explosion in California” events belong to the topic “Cell-phone Explosion”. Similarly, events such as “Theresa May is in Office”, “Theresa May’s Resignation Speech” can be covered by the topic “Brexit Negotiation”.

Furthermore, we define three types of edges, i.e., relationships, between nodes:

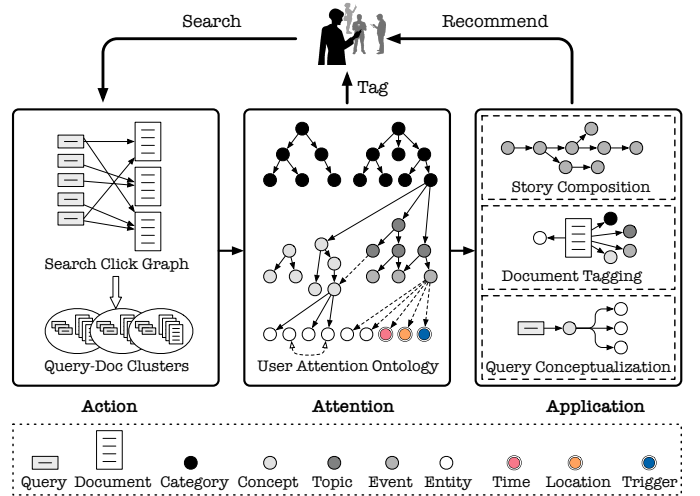
- **isA relationship**, indicating that the destination node is an instance of the source node. For example, the entity “Huawei Mate20 Pro” isAn instance of “Huawei Cellphones”.
- **involve relationship**, indicating that the destination node is involved in an event/topic described by the source node.
- **correlate relationship**, indicating two nodes are highly correlated with each other.

The edges in the Attention Ontology reveal the types of relationships and the degrees of relatedness between nodes. A plethora of edges enables the inference of more hidden interests of a user beyond the content he/she has browsed by moving along the edges on the Attention Ontology and recommending other related nodes at a coarser or finer granularity based on the nodes the user has visited. Furthermore, by analyzing edges between event nodes, we could also keep track of a *developing story*, which usually consists of a series of events, and keep interested users updated.

### 3 ONTOLOGY CONSTRUCTION

GIANT is a mechanism to discover phrases that users pay attention to from the web as well as to build a structured hierarchy out of them. In this section, we present our detailed techniques to construct the Attention Ontology based on neural networks and other ad-hoc methods. The entire process consists of two phases: i) user attention phrases mining, and ii) attention phrases linking. In the first phase, we define and extract user attention phrases in different semantic granularities from a large-scale search click graph. In the second phase, we link different extracted nodes and identify their relationships to construct a structured ontology.

Figure 2 shows the overall framework of GIANT, which constructs the Attention Ontology based on user search and click logs. The framework consists of three major components: action, attention, and application. In the action component, when users perform different actions (e.g., search, click, etc.), the user queries and clicked documents form a bipartite graph [66], commonly known as a search click graph. Based on it, we can collect highly correlated queries and documents



**Figure 2: Overview of our framework for constructing the Attention Ontology and performing different tasks.**

by aggregating documents that correspond to a query or vice versa, into query-doc clusters. In the attention component, we can extract different types of nodes (e.g., concepts, events, topics, entities, etc.) from the query-doc clusters, as well as learn the relationships between different nodes to form the Attention Ontology. In the application component, we can apply the Attention Ontology to different applications such as query conceptualization and document tagging. On the other hand, we can also integrate different nodes to user profiles to characterize the interest of different users based on his/her historical viewing behavior. In this manner, we can characterize both users and documents based on the Attention Ontology, which enables us to better understand and recommend documents from users’ perspective.

#### 3.1 Mining User Attentions

We propose a novel algorithm to extract various types of attention phrases (e.g., concepts, topics or events), which represent user attentions or interests, from a collection of queries and document titles.

**Problem Definition.** Suppose a bipartite search click graph  $G_{sc} = (Q, D, E)$  records the click-through information over a set of queries  $Q = \{q_1, q_2, \dots, q_{|Q|}\}$  and documents  $D = \{d_1, d_2, \dots, d_{|D|}\}$ . We use  $|\cdot|$  to denote the length of  $\cdot$ .  $E$  is a set of edges linking queries and documents. Our objective is to extract a set of phrases  $P = \{p_1, p_2, \dots, p_{|P|}\}$  from  $Q$  and  $D$  to represent user interests. Specifically, suppose  $p$  consists of a sequence of words  $\{w_{p1}, w_{p2}, \dots, w_{p|p|}\}$ . In our work, each phrase  $p$  is extracted from a subset of queries and the titles of correlated documents, and each word  $w_p \in p$  is contained by at least one query or title.





“seq” edge, indicating their order in the input. In Figure 3, the inverse direction of a “seq” edge points to the preceding words, which is indicated by a hollow triangle pointer. If the pair of nodes are not adjacent to each other, but there exists syntactical dependency between them, they will be connected by a bi-directional dashed edge which indicates the type of syntactical dependency relationship and the direction of it, while the inverse direction is also indicated by a hollow triangle pointer.

---

**ALGORITHM 2:** Constructing the Query-Title Interaction Graph

---

**Input:** a sequence of queries  $Q = \{q_1, q_2, \dots, q_{|Q|}\}$ , document titles  $T = \{t_1, t_2, \dots, t_{|T|}\}$ .

**Output:** a Query-Title Interaction Graph  $G_{qt}(Q, T)$ .

- 1: Create node set  $V = \{sos, eos\}$ , edge set  $E = \{\}$ ;
  - 2: **for** each input text passage  $x \in Q$  or  $x \in T$  **do**
  - 3:   append “sos” and “eos” as the first and the last token of  $x$ ;
  - 4:   construct a new node for each token in  $x$ ;
  - 5:   construct a bi-directional “seq” edge for each pair of adjacent tokens;
  - 6:   append each constructed node and edge into  $V$  or  $E$  only if the node is not contained by  $V$ , or no edge with the same source and target tokens exists in  $E$ ;
  - 7: **end for**
  - 8: **for** each input text passage  $x \in Q$  or  $x \in T$  **do**
  - 9:   perform syntactic parsing over  $x$ ;
  - 10:   construct a bi-directional edge for each dependency relationship;
  - 11:   append each constructed edge into  $E$  if no edge with the same source and target tokens exists in  $E$ ;
  - 12: **end for**
  - 13: construct graph  $G_{qt}(Q, T)$  from node set  $V$  and edge set  $E$ .
- 

Algorithm 2 shows the process of constructing a query-title interaction graph. We construct the nodes and edges by reading the inputs in  $Q_q$  and  $T_q$  in order. When we construct the edges between two nodes, as two nodes may have multiple adjacent relationships or syntactical relationships in different inputs, we only keep the first edge constructed. In this way, each pair of related nodes will only be connected by a bi-directional sequential edge or a syntactical edge. The idea is that we prefer the “seq” relationship as it shows a stronger connection than any syntactical dependency, and we prefer the syntactical relationships contained in the higher-weighted input text instead of the relationships in lower-weighted inputs. Compared with including all possible edges in a query-title interaction graph, empirical evidence suggests that our graph construction approach gives better performance for phrase mining.

After constructing a graph  $G_{qt}(Q_q, T_q)$  from a query-title cluster, we extract a phrase  $p$  by our GCTSP-Net, which contains a classifier to predict whether a node belong to  $p$ ,

and an asymmetric traveling salesman decoder to order the predicted positive nodes.

**Node Classification with R-GCN.** In the GCTSP-Net, we apply Relational Graph Convolutional Networks (R-GCN) [19, 27, 56] to our constructed QTIG to perform node classification.

Denote a directed and labeled multi-graph as  $G = (V, E, R)$  with labeled edges  $e_{vw} = (v, r, w) \in E$ , where  $v, w \in V$  are connected nodes, and  $r \in R$  is a relation type. A class of graph convolutional networks can be understood as a message-passing framework [19], where the hidden states  $h_v^l$  of each node  $v \in G$  at layer  $l$  are updated based on messages  $m_v^{l+1}$  according to:

$$m_v^{l+1} = \sum_{w \in N(v)} M_l(h_v^l, h_w^l, e_{vw}), \quad (3)$$

$$h_v^{l+1} = U_l(h_v^l, m_v^{l+1}), \quad (4)$$

where  $N(v)$  denotes the neighbors of  $v$  in graph  $G$ ,  $M_l$  is the message function at layer  $l$ , and  $U_l$  is the vertex update function at layer  $l$ .

Specifically, the message passing function of Relational Graph Convolutional Networks is defined as:

$$h_v^{l+1} = \sigma \left( \sum_{r \in R} \sum_{w \in N^r(v)} \frac{1}{c_{vw}} W_r^l h_w^l + W_0^l h_v^l \right), \quad (5)$$

where  $\sigma(\cdot)$  is an element-wise activation function such as  $\text{ReLU}(\cdot) = \max(0, \cdot)$ .  $N^r(v)$  is the set of neighbors under relation  $r \in R$ .  $c_{vw}$  is a problem-specific normalization constant that can be learned or pre-defined (e.g.,  $c_{vw} = |N^r(v)|$ ).  $W_r^l$  and  $W_0^l$  are learned weight matrices.

We can see that an R-GCN accumulates transformed feature vectors of neighboring nodes through a normalized sum. Besides, it learns relation-specific transformation matrices to take the type and direction of each edge into account. In addition, it adds a single self-connection of a special relation type to each node to ensure that the representation of a node at layer  $l + 1$  can also be informed by its representation at layer  $l$ .

To avoid the rapid growth of the number of parameters when increasing the number of relations  $|R|$ , R-GCN exploits basis decomposition and block-diagonal decomposition to regularize the weights of each layer. For basis decomposition, each weight matrix  $W_r^l \in \mathbb{R}^{d^{l+1} \times d^l}$  is decomposed as:

$$W_r^l = \sum_{b=1}^B a_{rb}^l V_b^l, \quad (6)$$

where  $V_b^l \in \mathbb{R}^{d^{l+1} \times d^l}$  are base weight matrices. In this way, only the coefficients  $a_{rb}^l$  depend on  $r$ . For block-diagonal decomposition,  $W_r^l$  is defined through the direct sum over a

set of low-dimensional matrices:

$$W_r^l = \bigoplus_{b=1}^B Q_{br}^l, \quad (7)$$

where  $W_r^l = \text{diag}(Q_{1r}^l, Q_{2r}^l, \dots, Q_{br}^l)$  is a block-diagonal matrix with  $Q_{br}^l \in \mathbb{R}^{(d^{l+1}/B) \times (d^l/B)}$ . The basis function decomposition introduces weight sharing between different relation types, while the block decomposition applies sparsity constraint on the weight matrices.

In our model, we apply R-GCN with basis decomposition to query-title interaction graphs to perform node classification. For each node in the graph, we represent it by a feature vector consisting of the embeddings of the token’s named entity recognition (NER) tag, part-of-speech (POS) tag, whether it is a stop word, number of characters in the token, as well as the sequential id that indicates the order each node be added to the graph. Using the concatenation of these embeddings as the initial node vectors, we pass the graph to a multi-layer R-GCN with a softmax( $\cdot$ ) activation (per node) on the output of the last layer. We label the nodes belonging to the target phrase  $p$  as 1 and other nodes as 0, and train the model by minimizing the binary cross-entropy loss on all nodes.

**Node Ordering with ATSP Decoding.** After classified a set of tokens  $V_p$  as target phrase nodes, the next step is to sort the nodes to get the final output. In our GCTSP-Net, we propose to model the problem as an asymmetric traveling salesman problem (ATSP), where the objective is to find the shortest route that starts from the “sos” node, visits each predicted positive nodes, and returns to the “eos” node. This approach is named as ATSP-decoding in our work.

We perform ATSP-decoding with a variant of the constructed query-title interaction graph. First, we remove all the syntactical dependency edges. Second, instead of connecting adjacent tokens by a bi-directional “seq” edge, we change it into unidirectional to indicate the order in input sequences. Third, we connect “sos” with the first predicted positive token in each input text, as well as connect the last predicted positive token in each input with the “eos” node. In this way, we remove the influence of prefixing and suffixing tokens in the inputs when finding the shortest path. Finally, the distance between a pair of predicted nodes is defined as the length of the shortest path in the modified query-title interaction graph. In this way, we can solve the problem with Lin-Kernighan Traveling Salesman Heuristic [23] to get a route of the predicted nodes and output  $p$ .

We shall note that ATSP-decoding will produce a phrase that contains only unique tokens. In our work, we observe that only less than 1% attention phrases contain duplicated tokens, while most of the duplicated tokens are punctuations. Even if we need to produce duplicate tokens when applying our model to other tasks, we just need to design task-specific

heuristics to recognize the potential tokens (such as count their frequency in each query and title), and construct multiple nodes for it in the query-title interaction graph.

**Attention Phrase Normalization.** The same user attention may be expressed by slightly different phrases. After extracting a phrase using GCTSP-Net, we merge highly similar phrases into one node in our Attention Ontology. Specifically, we examine whether a new phrase  $p_n$  is similar to an existing phrase  $p_e$  by two criteria: i) the non-stop words in  $p_n$  shall be similar (same or synonyms) with that in  $p_e$ , and ii) the TF-IDF similarity between their context-enriched representations shall be above a threshold  $\delta_m$ . The context-enriched representation of a phrase is obtained by using itself as a query and concatenating the top 5 clicked titles.

**Training Dataset Construction.** To reduce human effort and accelerate the labeling process of training dataset creation, we design effective unsupervised strategies to extract candidate phrases from queries and titles, and provide the extracted candidates together with query-title clusters to workers as assistance. For concepts, we combine bootstrapping with query-title alignment [34]. The bootstrapping strategy exploits pattern-concept duality: we can extract a set of concepts from queries following a set of patterns, and we can learn a set of new patterns from a set of queries with extracted concepts. Thus, we can start from a set of seed patterns, and iteratively accumulate more and more patterns and concepts. The query-title alignment strategy is inspired by the observation that a concept in a query is usually mentioned in the clicked titles associated with the query, yet possibly in a more detailed manner. Based on this observation, we align a query with its top clicked titles to find a title chunk which fully contains the query tokens in the same order and potentially contains extra tokens within its span. Such a title chunk is selected as a candidate concept.

For events, we split the original unsegmented document titles into subtitles by punctuations and spaces. After that, we only keep the set of subtitles with lengths between  $L_l$  (we use 6) and  $L_h$  (we use 20). For each remaining subtitle, we score it by counting how many unique non-stop query tokens within it. The subtitles with the same score will be sorted by its click-through rate. Finally, we select the top ranked subtitle as a candidate event phrase.

**Attention Derivation.** After extracting a collection of attention nodes (or phrases), we can further derive higher level concepts or topics from them, which automatically become their parent nodes in our Attention Ontology.

On one hand, we derive higher-level concepts by applying Common Suffix Discovery (CSD) to extracted concepts. We perform word segmentation over all concept phrases, and find out the high-frequency suffix words or phrases. If the suffixes forms a noun phrase, we add it as a new concept node. For example, the concept “animated film (动画|电影)”

can be derived from “famous animated film (著名的|动画|电影)”, “award-winning animated film (获奖的|动画|电影)” and “Hayao Miyazaki animated film (宫崎骏|动画|电影)”, as they share the common suffix “animated film (动画|电影)”

On the other hand, we drive high-level topics by applying Common Pattern Discovery (CPD) to extracted events. We perform word segmentation, named entity recognition and part-of-speech tagging over the event phrases. Then we find out high-frequency event patterns and recognize the different elements in the events. If the elements (e.g., entities or locations of events) have *isA* relationship with one or multiple common concepts, we replace the different elements by the most fine-grained common concept ancestor in the ontology. For example, we can derive a topic “Singer will have a concert (歌手|开|演唱会)” from “Jay Chou will have a concert (周杰伦|开|演唱会)” and “Taylor Swift will have a concert (泰勒|斯威夫特|开|演唱会)”, as the two phrases sharing the same pattern “XXX will have a concert (XXX|开|演唱会)”, and both “Jay Chou (周杰伦)” and “Taylor Swift (泰勒|斯威夫特)” belong to the concept “Singer (歌手)”. To ensure that the derived topic phrases are user interests, we filter out phrases that have not been searched by a certain number of users.

### 3.2 Linking User Attention

The previous step produces a large set of nodes representing user attentions (or interests) in different granularities. Our goal is to construct a taxonomy based on these individual nodes to show their correlations. With edges between different user attentions, we can reason over it to infer a user’s real interest.

In this section, we describe our methods to link attention nodes and construct the complete ontology. We will construct the *isA* relationships, *involve* relationships, and *correlate* relationships between categories, extracted attention nodes and entities to construct an ontology. We exploit the following action-driven strategies to link different types of nodes.

**Edges between Attention and Categories.** To identify the *isA* relationship between *attention-category* pairs, we utilize the co-occurrence of them shown in user search click logs. Given an attention phrase  $p$  as the search query, suppose there are  $n^p$  clicked documents of search query  $p$  in the search click logs, and among them there are  $n_g^p$  documents belong to category  $g$ . We then estimate  $\mathbb{P}(g|p)$  by  $\mathbb{P}(g|p) = n_g^p/n^p$ . We identify that there is an *isA* relationship between  $p$  and  $g$  if  $\mathbb{P}(g|p) > \delta_g$  (we set  $\delta_g = 0.3$ ).

**Edges between Attention.** To discover *isA* relationships, we utilize the same criteria when we perform attention derivation: we link two concepts by the *isA* relationship if one concept is the suffix of another concept, and we link two

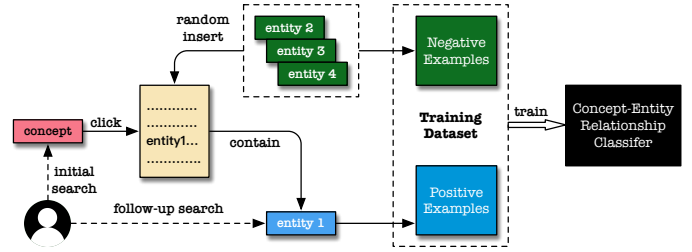


Figure 4: Automatic construction of the training datasets for classifying the *isA* relationship between concepts and entities.

topic/event attentions by the *isA* relationship if they share the same pattern and there exists *isA* relationships between their non-overlapping tokens. Note that if a topic/event doesn’t contain an element of another topic/event phrase, it also indicates that they have *isA* relationship. For example, “Jay Chou will have a concert” has *isA* relationship with both “Singer will have a concert” and “have a concert”. For the *involve* relationship, we connect a concept to a topic if the concept is contained in the topic phrase.

**Edges between Attention and Entities.** We extract: i) *isA* relationships between concepts and entities; ii) *involve* relationships between topics/events and entities, locations or triggers.

For concepts and entities, using strategies such as co-occurrence will introduce a lot of noises, as co-occurrence doesn’t always indicate an entity belongs to a concept. To solve this issue, we propose to train a concept-entity relationship classifier based on the concept and the entity’s context information in the clicked documents. Labeling a training dataset for this task requires a lot of human efforts. Instead of manual labeling, we propose a method to automatically construct a training dataset from user search click graphs. Figure 4 shows how we construct such a training dataset. We select the concept-entity pairs from search logs as positive examples if: i) the concept and the entity are two consecutive queries from one user, and ii) the entity is mentioned by a document which a user clicked after issuing the concept as a search query. Besides, we select entities belonging to the same higher-level concept or category, and insert them into random positions of the document to create negative examples of the dataset. For the classifier, we can train a classifier such as GBDT based on manual features, or fine-tune a pre-trained language model to incorporate semantic features and infer whether the context indicates a *isA* relationship between the concept-entity pair.

For events/topics and entities, we only recognize the important entities, triggers and locations in the event/topic, and connect them by an *involve* relationship edge. We first



create an initial dataset by extracting all the entities, locations, and trigger words in the events/topic based on a set of predefined trigger words and entities. Then the dataset is manually revised by workers to remove the unimportant elements. Based on this dataset, we reuse our GCTSP-Net and train it without ATSP-decoding to perform 4-class (entity, location, trigger, other) node classification over the query-title interaction graphs of the events/topics. In this way, we can recognize the different elements of an event or topic, and construct *involve* edges between them.

**Edges between Entities.** Finally, we construct the *correlate* relationship between entity pairs by the co-occurrence information in user queries and documents. We utilize high frequency co-occurring entity pairs in queries and documents as positive entity pairs, and perform negative sampling to create negative entity pairs. After automatically created a dataset from search click logs and web documents, we learn the embedding vectors of entities with Hinge loss, so that the Euclidean distance between two correlated entities will be small. After learned the embedding vectors of different entities, we classify a pair of entities as correlated if their Euclidean distance is smaller than a threshold.

Note that the same approach for *correlate* relationship discovery can be applied to other type of nodes such as concepts. Currently, we only constructed such relationships between entities.

## 4 APPLICATIONS

In this section, we show how our attention ontology can be applied to a series of NLP tasks to achieve user-centered text understanding.

**Story Tree Formation.** The relationships between events and the involved entities, triggers and locations can be utilized to cluster correlated events and form a story tree [35]. A story tree organizes a collection of related events with a tree structure to highlight the evolving nature of a real-world story. Given an event  $p^e$ , we retrieve a collection of related events  $P^e$  and use them to form a story tree, which allows us to better serve the users by recommending follow-up events from  $P^e$  when they have read news articles about  $p^e$ .

Constructing a story tree from an attention ontology involves four steps: retrieving correlated events, calculating similarity matrix, hierarchical clustering, and tree formation. First, with the help of the attention ontology, we can retrieve related events set  $P^e$  give an event  $p^e$ . Specifically, the criteria to retrieve “correlated” events can be flexible. For example, we can set a requirement that each event  $p_i \in P^e$  shares at least one common child entity with  $p^e$ , or we can force the triggers of them to be the same. Second, we can estimate the similarities between each pair of events based on the text similarity of event phrases and the common entities, triggers

or locations shared by them. Specifically, we calculate the similarity between two events by:

$$s(p_1^e, p_2^e) = f_m(p_1^e, p_2^e) + f_g(p_1^e, p_2^e) + f_e(p_1^e, p_2^e), \quad (8)$$

$$f_m(p_1^e, p_2^e) = \text{CosineSimilarity}(\mathbf{v}^{p_1^e}, \mathbf{v}^{p_2^e}), \quad (9)$$

$$f_g(p_1^e, p_2^e) = \text{CosineSimilarity}(\mathbf{v}^{g_1^e}, \mathbf{v}^{g_2^e}), \quad (10)$$

$$f_e(p_1^e, p_2^e) = \text{TF-IDFSimilarity}(E^{p_1^e}, E^{p_2^e}), \quad (11)$$

where  $s(p_1^e, p_2^e)$  is the measured similarity between the two events, It is given by the sum of three scores: i)  $f_m(p_1^e, p_2^e)$ , which represents the semantic distance between two event phrases. We use the cosine similarity of BERT-based phrase encoding vectors  $\mathbf{v}^{p_1^e}$  and  $\mathbf{v}^{p_2^e}$  for the two events [14]; ii)  $f_g(p_1^e, p_2^e)$ , which represents the similarity of the triggers in two events. We calculate the similarity between trigger  $g_1^e$  in event  $p_1^e$  and  $g_2^e$  in  $p_2^e$  by the cosine similarity of the word vectors  $\mathbf{v}^{g_1^e}$  and  $\mathbf{v}^{g_2^e}$  from [61]; iii)  $f_e(p_1^e, p_2^e)$ , the TF-IDF similarity between the set of entities  $E^{p_1^e}$  of event  $p_1^e$  and  $E^{p_2^e}$  of  $p_2^e$ . After the measurement of the similarities between events, we perform hierarchical clustering to group them into hierarchical clusters. Finally, we order the events by time, and put the events in the same cluster into the same branch. In this way, the cluster hierarchy is transformed into the branches of a tree.

**Document Tagging** We can also utilize the attention phrases to describe the main topics of a document by tagging the correlated attentions to the document, even if the phrase is not explicitly mentioned in the document. For example, a document talking about films “Captain America: The First Avenger”, “Avengers: Endgame” and “Iron Man” can be tagged with the concept “Marvel Super Hero Movies” even though the concept may not be contained by it. Similarly, a document talking about “Theresa May’s Resignation Speech” can be tagged by topics “Brexit Negotiation”, while traditional keyword-based methods are not able to reveal such correlations.

To tag concepts to documents, we combine a matching-based approach and a probabilistic inference-based approach based on the key entities in a document. Suppose  $d$  contains a set of key entities  $E^d$ . For each entity  $e^d \in E^d$ , we obtain its parent concepts  $P^c$  in the attention ontology as candidate tags. For each candidate concept  $p^c \in P^c$ , we score the coherence between  $d$  and  $p^c$  by calculating the TF-IDF similarity between the title of  $d$  and the context-enriched representation of  $p^c$  (i.e., the topic clicked titles of  $p^c$ ).

When no parent concept can be found by the attention ontology, we identify relevant concepts by utilizing the context information of the entities in  $d$ . Denote the probability that concept  $p^c$  is related to document  $d$  as  $\mathbb{P}(p^c|d)$ . We estimate

it by:

$$\mathbb{P}(p^c|d) = \sum_{i=1}^{|E^d|} \mathbb{P}(p^c|e_i^d)\mathbb{P}(e_i^d|d), \quad (12)$$

$$\mathbb{P}(p^c|e_i^d) = \sum_{j=1}^{|X_{e_i^d}|} \mathbb{P}(p^c|x_j)\mathbb{P}(x_j|e_i^d), \quad (13)$$

$$\mathbb{P}(p^c|x_j) = \begin{cases} \frac{1}{|P_{x_j}^c|} & \text{if } x_j \text{ is a substring of } p^c, \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

where  $E^d$  is the key entities of  $d$ ,  $\mathbb{P}(e_i^d|d)$  is the document frequency of entity  $e_i^d \in E^d$ .  $\mathbb{P}(p^c|e_i^d)$  estimates the probability of concept  $p^c$  given  $e_i^d$ , which is inferred from the context words of  $e_i^d$ .  $\mathbb{P}(x_j|e_i^d)$  is the co-occurrence probability of context word  $x_j$  with  $e_i^d$ . We consider two words as co-occurred if they are contained in the same sentence.  $X_{e_i^d}$  are the set of contextual words of  $e_i^d$  in  $d$ .  $P_{x_j}^c$  is the set of concepts containing  $x_j$  as a substring.

To tag events or topics to a document, we combine longest common subsequence based (LCS-based) textural matching with Duet-based semantic matching [42]. For LCS-based matching, we concatenate a document title with the first sentence in content, and calculate the length of longest common subsequence between a topic/event phrase and the concatenated string. For Duet-based matching, we utilize the Duet neural network [42] to classify whether the phrase matches with the concatenated string. If the length of the longest common subsequence is above a threshold and the classification result is positive, we tag the phrase to the document.

**Query Understanding.** A user used to search about an entity may be interested in a broader class of similar entities. However, the user may not be even aware of the entities similar to the query. With the help of our ontology, we can better understand users’ implicit intention and perform query conceptualization or recommendation to improve the user experience in search engines. Specifically, we analyze whether a query  $q$  contains a concept  $p^c$  or an entity  $e$ . If a query conveys a concept  $p^c$ , we can rewrite it by concatenating  $q$  with each of the entities  $e_i$  that have *isA* relationship with  $p^c$ . In this way, we rewrite the query to the format of “ $q e_i$ ”. If a query conveys an entity  $e$ , we can perform query recommendation by recommend users the entities  $e_i$  that have *correlate* relationship with  $e$  in the ontology.

## 5 EVALUATION

Our proposed GIANT system is the core ontology system in multiple applications in Tencent, i.e., Tencent QQ Browser, Mobile QQ and WeChat, and is serving more than a billion daily active users all around the world. It is implemented by Python 2.7 and C++. Each component of our system works as

	Category	Concept	Topic	Event	Entity
Quantity	1, 206	460, 652	12, 679	86, 253	1, 980, 841
Grow / day	-	11, 000	-	120	-

**Table 1: Nodes in the attention ontology.**

	<i>isA</i>	<i>correlate</i>	<i>involve</i>
Quantity	490, 741	1, 080, 344	160, 485
Accuracy	95%+	95%+	99%+

**Table 2: Edges in the attention ontology.**

a service and is deployed with Tars<sup>1</sup>, a high-performance remote procedure call (RPC) framework based on name service and Tars protocol. The attention mining and linking services are deployed on 10 dockers, with each configured with four processor Intel(R) Xeon(R) Gold 6133 CPU @ 2.50GHz and 6GB memory. Applications such as document tagging are running on 48 dockers with the same configuration. MySQL database is used for data storage. We construct the attention ontology from large-scale real-world daily user search click logs. While our system is deployed on Chinese datasets, the techniques proposed in our work can be easily adapted to other languages.

### 5.1 Evaluation of the Attention Ontology

Table 1 shows the statistics of different nodes in the attention ontology. We extract attention phrases from daily user search click logs. Therefore, the scale of our ontology keeps growing. Currently, our ontology contains 1, 206 predefined categories, 460, 652 concepts, 12, 679 topics, 86, 253 events and 1, 980, 841 entities. We are able to extract around 27, 000 concepts and 400 events every day, and around 11, 000 concepts and 120 events are new. For online concept and event tagging, our system processes 350 documents per second.

Table 2 shows the statistics and accuracies of different types of edges (relationships) in the attention ontology. Currently, we have constructed more than 490K *isA* relationships, 1, 080K *correlate* relationships, and 160K *involve* relationships between different nodes. The human evaluation performed by three managers in Tencent shows the accuracies of the three relationship types are above 95%, 95% and 99%, respectively.

To give intuition into what kind of attention phrases can be derived from search click graphs, Table 3 and Table 4 show a few typical examples of concepts and events (translated from Chinese), and some topics, categories, and entities that share *isA* relationship with them. By fully exploiting the information of user actions contained in search click graphs, we transform user actions to user attentions, and extract

<sup>1</sup><https://github.com/TarsCloud/Tars>

Categories	Concepts	Instances
Sports	Famous long-distance runner	Dennis Kipruto Kimetto, Kenenisa Bekele
Stars	Actors who committed suicide	Robin Williams, Zhang Guorong, David Strickland
Drama series	American crime drama series	American Crime Story, Breaking Bad, Criminal Minds
Fiction	Detective fiction	Adventure of Sherlock Holmes, The Maltese Falcon

**Table 3: Showcases of concepts and the related categories and entities.**

Categories	Topics	Events	Entities
Music	Singers win music awards	Jay Chou won the Golden Melody Awards in 2002, Taylor Swift won the 2016 Grammy Awards	Jay Chou, Taylor Swift
Cellphone	cellphone launch events	Apple news conferences 2018 mid-season, Samsung Galaxy S9 officially released	Apple, iPhone, Samsung, Galaxy S9
Esports	League of Legends season 8	LOL S8 finals announcement, IG wins the S8 final, IG’s reward for winning the S8 final revealed	League of Legends, IG team, finals

**Table 4: Showcases of events and the related categories, topics and involved entities.**

concepts such as “Actors who committed suicide (自杀的演员)”. Besides, we can also extract events or higher level topics of users’ interests, such as “Taylor Swift and Katy Perry” if a user often inputs related queries. Based on the connections between entities, concepts, events and topics, we can also infer what a user really cares and improve the performance of recommendation.

## 5.2 Evaluation of the GCTSP-Net

We evaluate our GCTSP-Net on multiple tasks by comparing it to a variety of baseline approaches.

**Datasets.** To the best of our knowledge, there is no publicly available dataset suitable for the task of heterogeneous phrase mining from user queries and search click graphs. To construct the user attention ontology, we create two large-scale datasets for concept mining and event mining using the approaches described in Sec. 3: the Concept Mining Dataset (CMD) and the Event Mining Dataset (EMD). These two datasets contain 10, 000 examples and 10, 668 examples respectively. Each example is composed of a set of correlated queries and top clicked document titles from real-world query logs, together with a manually labeled gold

phrase (concept or event). For the event mining dataset, it further contains triggers, key entities and locations of each event. We use the earliest article publication time as the time of each event example. The datasets are labeled by 3 professional product managers in Tencent and 3 graduate students. For each dataset, we utilize 80% as training set, 10% as development set, and 10% as test set. The datasets will be published for research purposes <sup>2</sup>.

**Methodology and Models for Comparison.** We compare our GCTSP-Net with the following baseline methods on the concept mining task:

- **TextRank.** A classical graph-based keyword extraction model [41].<sup>3</sup>
- **AutoPhrase.** A state-of-the-art phrase mining algorithm that extracts quality phrases based on POS-guided segmentation and knowledge base [59].<sup>4</sup>
- **Match.** Extract concepts from queries and titles by matching using patterns from bootstrapping [34].
- **Align.** Extract concepts by the query-title alignment strategy described in Sec. 3.1.
- **MatchAlign.** Extract concepts by both pattern matching and query-title alignment strategy.
- **LSTM-CRF-Q.** Apply LSTM-CRF [25] to input query.
- **LSTM-CRF-T.** Apply LSTM-CRF [25] to titles.

For the TextRank and AutoPhrase algorithm, we extract the top 5 keywords or phrases from queries and titles, and concatenate them in the same order with the query/title to get the extracted phrase. For MatchAlign, we select the most frequent result if multiple phrases are extracted. For LSTM-CRF-Q/LSTM-CRF-T, it consists of a 200-dimensional word embedding layer initialized with the word vectors proposed in [61], a BiLSTM layer with hidden size 25 for each direction, and a Conditional Random Field (CRF) layer which predicts whether each word belongs to the output phrase by Beginning-Inside-Outside (BIO) tags.

For event mining task, we compare with TextRank and LSTM-CRF. In addition, we also compare with:

- **TextSummary [41].** An encoder-decoder model with attention mechanism for text summarization.<sup>5</sup>
- **CoverRank.** Rank queries and subtitles by counting the covered nonstop query words, as described in 3.1.

For TextRank, we use the top 2 queries and top 2 selected subtitles given by CoverRank, and perform re-ranking. For TextSummary, we use the 200-dimensional word embeddings in [61], two-layer BiLSTM (256 hidden size for each direction) as encoder, and one layer LSTM with 512 hidden size and attention mechanism as decoder (beam size for decoding

<sup>2</sup><https://github.com/BangLiu/GIANT>

<sup>3</sup><https://github.com/letiantian/TextRank4ZH>

<sup>4</sup><https://github.com/shangjingbo1226/AutoPhrase>

<sup>5</sup><https://github.com/dongjun-Lee/text-summarization-tensorflow>

Method	EM	F1	COV
TextRank	0.1941	0.7356	1
AutoPhrase	0.0725	0.4839	0.9353
Match	0.1494	0.3054	0.3639
Align	0.7016	0.8895	0.9611
MatchAlign	0.6462	0.8814	0.9700
Q-LSTM-CRF	0.7171	0.8828	0.9731
T-LSTM-CRF	0.3106	0.6333	0.9062
GCTSP-Net	<b>0.783</b>	<b>0.9576</b>	<b>1</b>

**Table 5: Compare concept mining approaches.**

Method	EM	F1	COV
TextRank	0.3968	0.8102	1
CoverRank	0.4663	0.8169	1
TextSummary	0.0047	0.1064	1
LSTM-CRF	0.4597	0.8469	1
GCTSP-Net	<b>0.5164</b>	<b>0.8562</b>	<b>0.9972</b>

**Table 6: Compare event mining approaches.**

is 10). We feed the concatenation of queries and titles into TextSummary to generate the output. For LSTM-CRF, the LSTM layer in it is configured similarly to the encoder of TextSummary. We feed each title individually into it to get different outputs, filter the outputs by length (number of characters between 6 and 20), and finally select the phrase which belongs to the top clicked title.

For event key elements recognition (key entities, trigger, location), it is a 4-class classification task over each word. We compare our model with LSTM and LSTM-CRF. The difference between LSTM and LSTM-CRF is that LSTM replaces the CRF layer in LSTM-CRF with a softmax layer.

For each baseline, we individually tune the hyper-parameters in order to achieve its best performance. As to our approach (GCTSP-Net), we stack 5-layer R-GCN with hidden size 32 and number of bases  $B = 5$  in basis decomposition for graph encoding and node classification. We will open-source our code together with the datasets for research purposes.

**Metrics.** We use Exact Match (EM), F1 and coverage rate (COV) to evaluate the performance of phrase mining tasks. The exact match score is 1 if the prediction is exactly the same as ground-truth or 0 otherwise. F1 measures the portion of overlap tokens between the predicted phrase and the ground-truth phrase [52]. The coverage rate measures the percentage of non-empty predictions of each approach. For the event key elements recognition task, we evaluate by the F1-macro, F1-micro, and F1-weighted metrics.

**Evaluation results and analysis.** Table 5, Table 6, and Table 7 compare our model with different baselines on the CMD and EMD datasets for concept mining, event mining, and event key elements recognition. We can see that our

Method	F1-macro	F1-micro	F1-weighted
LSTM	0.2108	0.5532	0.6563
LSTM-CRF	0.2610	0.6468	0.7238
GCTSP-Net	<b>0.6291</b>	<b>0.9438</b>	<b>0.9331</b>

**Table 7: Compare event key elements recognition approaches.**

unified model for different tasks significantly outperforms all baseline approaches. The outstanding performance can be attribute to: first, our query-title interaction graph efficiently encodes the information of word adjacency, word features, dependencies, query-title overlapping and so on in a structured manner, which are critical for both attention phrase mining tasks and event key elements recognition tasks. Second, the multi-layer R-GCN encoder in our model can learn from both the node features and the multi-resolution structural patterns from query-title interaction graph. Therefore, combining query-title interaction graph with R-GCN encoder, we can achieve great performance in different node classification tasks. Furthermore, the unsupervised ATSP-decoding sorts the extracted tokens to form an ordered phrase efficiently. In contrast, heuristic-based approaches and LSTM-based approaches are not robust to the noises in dataset, and cannot capture the structural information in queries and titles. In addition, existing keyphrase extraction methods such as TextRank [41] and AutoPhrase [59] are better suited for extracting key words or phrases from long documents, lacking the ability to give satisfactory performance in our tasks.

### 5.3 Applications: Story Tree Formation and Document Tagging

**Story Tree Formation.** We apply the story tree formation algorithm described in Sec. 4 to real-world events to test its performance. Figure 5 gives an example to illustrate what we can obtain through our approach. In the example, each node is an event, together with the documents that can be tagged by this event. We can see that our method can successfully cluster events related to “China-US Trade”, ordering the events by the published time of the articles, and show the evolving structure of coherent events. For example, the branch consists of events 3 ~ 6 are mainly about “Sino-US trade war is emerging”, the branch 8 ~ 10 are resolving around “US agrees limited trade deal with China”, and 11 ~ 14 are about “Impact of Sino-US trade war felt in both countries”. By clustering and organizing events and related documents in such a tree structure, we can track the development of different stories (clusters of coherent events),

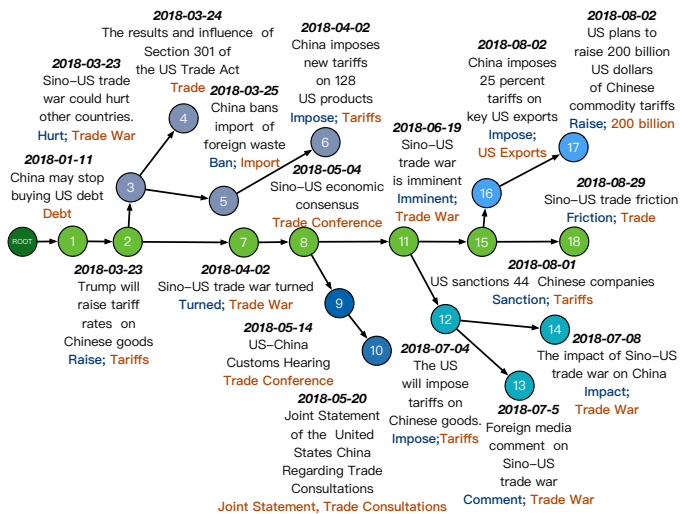


Figure 5: An example to show the constructed story tree given by our approach.

reduce information redundancy, and improve document recommendation by recommending users the follow-up events they are interested in.

**Document Tagging.** For document tagging, our system currently processes around 1,525,682 documents per day, where about 35% of them can be tagged with at least one concept, and 4% can be tagged with an event. We perform human evaluation by sampling 500 documents for each major category (“game”, “technology”, “car”, and “entertainment”). The result shows that the precision of concept tagging for documents is 88% for “game” category, 91% for “technology”, 90% for “car”, and 87% for “entertainment”. The overall precision for documents of all categories is 88%. As to event tagging on documents, the overall precision is 96%.

## 5.4 Online Recommendation Performance

We evaluate the effect of our attention ontology on recommendations by analyzing its performance in the news feeds stream of Tencent QQ Browser which has more than 110 million daily active users. In the application, both users and articles are tagged with categories, topics, concepts, events or entities from the attention ontology, as shown in Figure 2. The application recommends news articles to users based on a variety of strategies, such as content-based recommendation, collaborative filtering and so on. For the content-based recommendation, it matches users with articles through the common tags they share.

We analyze the Click-Through Rate (CTR) given by different tags from July 16, 2019 to August 15, 2019 to evaluate their effects. Click-through rate is the ratio of the number of users who click on a recommended link to the number of total users who received the recommendation. Figure 6

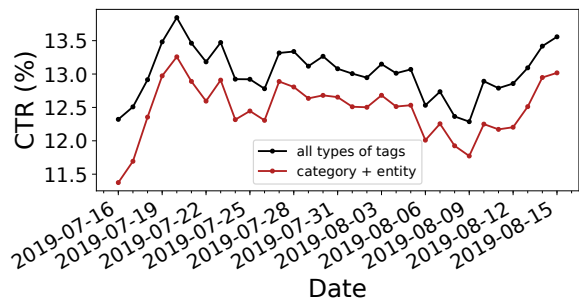


Figure 6: The click-through rates with/without extracted tags.

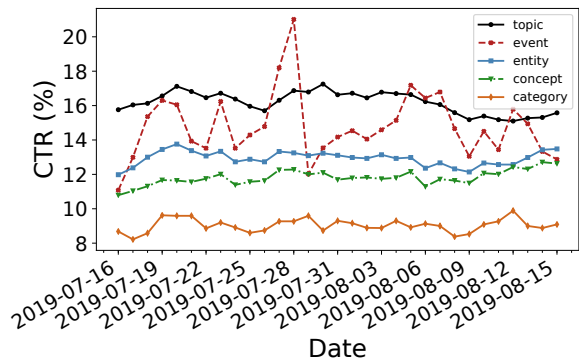


Figure 7: The click-through rates of different tags.

compares the CTR when we recommend documents to users with different strategies. Traditional recommender systems only utilize the category tags and entity tags. We can see that including more types of attention tags (topics, events, or concepts) in recommendation can constantly help improving the CTR on different days: the average CTR improved from 12.47% to 13.02%. The reason is that the extracted concepts, events or topics can depict user interests with suitable granularity and help to solve the inaccurate recommendation and monotonous recommendation problems.

We further analyze the effects of each type of tags in recommendation. Figure 7 shows the CTR of the recommendations given by different types of tags. The average CTR for topic, event, entity, concept and category are 16.18%, 14.78%, 12.93%, 11.82%, and 9.04%, respectively. The CTR of both topic tags and event tags are much higher than that of category and entities, which shows the effectiveness of our constructed ontology. For events, the events happening on each days dramatically different with each other, and they are not always attractive to users. Therefore, the CTR of event tags is less stable than the topic tags. For concept tags, they are generalization of fine-grained entities and has *isA* relationship with them. As there may have noises when we perform user interest inference using the relationships between entities and concepts, the CTR of concepts are slightly



lower than entities. However, compared to entities, our experience show that concepts can significantly increase the diversity of recommendation and are helpful in solving the problem of monotonous recommendation.

## 6 RELATED WORK

Our work is mainly related to the following research lines.

**Taxonomy and Knowledge Base Construction.** Most existing taxonomy or knowledge bases, such as Probase [69], DBpedia [32], YAGO [63], extract general concepts about the world and construct graphs or taxonomies based on Wikipedia or formal documents. In contrast, our work utilizes search click graphs to construct an ontology for describing user interests or attentions. Our prior work [34] constructs a three-layered taxonomy from search logs. Compared to it, our GIANT system constructs an attention ontology with more types of nodes and relationships based a novel algorithm for heterogeneous phrase mining. There are also works that construct a taxonomy from keywords [39] or queries [5]. Biperpedia [22] extracts class-attribute pairs from query stream to expand a knowledge graph. [49] extracts classes of instances with attributes and class labels from web documents and query logs.

**Concept Mining.** Existing approaches on concept mining are closely related to research works on named entity recognition [31, 44, 54], term recognition [18, 48, 74], keyphrase extraction [16, 68] or quality phrase mining [34, 36, 59]. Traditional algorithms utilize pre-defined part-of-speech (POS) templates and dependency parsing to identify noun phrases as term candidates [29, 59]. Supervised noun phrase chunking techniques [9, 51] automatically learn rules for identifying noun phrase boundaries. There are also methods that utilize resources such as knowledge graph to further enhance the precision [53, 67]. Data-driven approaches make use of frequency statistics in the corpus to generate candidate terms and evaluate their quality [16, 36, 47]. Phrase quality-based approaches exploit statistical features to measure phrase quality, and learn a quality scoring function by using knowledge base entity names as training labels [36, 59]. Neural network-based approaches consider the problem as sequence tagging and train complex deep neural models based on CNN or LSTM-CRF [25].

**Event Extraction.** Existing research works on event extraction aim to identify different types of event triggers and their arguments from unstructured text data. They combine supervised or semi-supervised learning with features derived from training data to classify event types, triggers and arguments [10, 24, 26, 37, 45]. However, these approaches cannot be applied to new types of events without additional annotation effort. The ACE2005 corpus [20] includes event annotations for 33 types of events. However, such small

hand-labeled data is hard to train a model to extract maybe thousands of event types in real-world scenarios. There are also works using neural networks such as RNNs [45, 58], CNNs [11, 46] or GCNs [38] to extract events from text. Open domain event extraction [55, 64] extracts news-worthy clusters of words, segments and frames from social media data such as Twitter [4], usually under unsupervised or semi-supervised settings and exploits information redundancy.

**Relation Extraction.** A comprehensive introduction about relation extraction can be found in [50]. Most existing techniques for relation extraction can be classified into the following classes. First, supervised learning techniques, such as features-based [21] and kernel based [13] approaches, require entity pairs that labeled with one of the pre-defined relation types as the training dataset. Second, semi-supervised approaches, including bootstrapping [7], active learning [33, 57] and label propagation [8], exploit the unlabeled data to reduce the manual efforts of creating large-scale labeled dataset. Third, unsupervised methods [71] utilize techniques such as clustering and named entity recognition to discover relationships between entities. Fourth, Open Information Extraction [17] construct comprehensive systems to automatically discover possible relations of interest using text corpus. Last, distant supervision based techniques leverage pre-existing structured or semi-structured data or knowledge to guide the extraction process [60, 73].

## 7 CONCLUSION

In this paper, we describe our design and implementation of GIANT, a system that proposed to construct a web-scale user attention ontology from large amount of query logs and search click graphs for various applications. It consists of around two millions of heterogeneous nodes with three types of relationships between them, and keeps growing with newly retrieved nodes and identified relationships every day. To construct the ontology, we propose the query-title interaction graph to represent the correlations (such as adjacency or syntactical dependency) between the tokens in correlated queries and document titles. Furthermore, we propose the GCTSP-Net to extract multi-type phrases from the query-title interaction graph, as well as recognize the key entities, triggers or locations in events. After constructing the attention ontology by our models, we apply it to different applications, including document tagging, story tree formation, as well as recommendation. We run extensive experiments and analysis to evaluate the quality of the constructed ontology and the performance of our new algorithms. Results show that our approach outperforms a variety of baseline methods on three tasks. In addition, our attention ontology significantly improves the CTR of news feeds recommendation in a real-world application.

## REFERENCES

- [1] Gediminas Adomavicius and Alexander Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering* 6 (2005), 734–749.
- [2] Gediminas Adomavicius and Alexander Tuzhilin. 2011. Context-aware recommender systems. In *Recommender systems handbook*. Springer, 217–253.
- [3] Chinatsu Aone and Mila Ramos-Santacruz. 2000. REES: a large-scale relation and event extraction system. In *Proceedings of the sixth conference on Applied natural language processing*. Association for Computational Linguistics, 76–83.
- [4] Farzindar Atefeh and Wael Khreich. 2015. A survey of techniques for event detection in twitter. *Computational Intelligence* 31, 1 (2015), 132–164.
- [5] Ricardo Baeza-Yates and Alessandro Tiberi. 2007. Extracting semantic relations from query logs. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. 76–85.
- [6] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. 2013. Recommender systems survey. *Knowledge-based systems* 46 (2013), 109–132.
- [7] Sergey Brin. 1998. Extracting patterns and relations from the world wide web. In *International Workshop on The World Wide Web and Databases*. Springer, 172–183.
- [8] Jinxiu Chen, Donghong Ji, Chew Lim Tan, and Zhengyu Niu. 2006. Relation extraction using label propagation based semi-supervised learning. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 129–136.
- [9] Kuang-hua Chen and Hsin-Hsi Chen. 1994. Extracting noun phrases from large-scale texts: A hybrid approach and its automatic evaluation. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 234–241.
- [10] Yubo Chen, Shulin Liu, Xiang Zhang, Kang Liu, and Jun Zhao. 2017. Automatically labeled data generation for large scale event extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 409–419.
- [11] Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 167–176.
- [12] Mário Cordeiro and João Gama. 2016. Online social networks event detection: a survey. In *Solving Large Scale Learning Tasks. Challenges and Algorithms*. Springer, 1–41.
- [13] Aron Culotta and Jeffrey Sorensen. 2004. Dependency tree kernels for relation extraction. In *Proceedings of the 42nd annual meeting on association for computational linguistics*. ACL, 423.
- [14] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [15] George R Doddington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie M Strassel, and Ralph M Weischedel. 2004. The Automatic Content Extraction (ACE) Program-Tasks, Data, and Evaluation. In *Lrec*, Vol. 2. Lisbon, 1.
- [16] Ahmed El-Kishky, Yanglei Song, Chi Wang, Clare R Voss, and Jiawei Han. 2014. Scalable topical phrase mining from text corpora. *Proceedings of the VLDB Endowment* 8, 3 (2014), 305–316.
- [17] Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the conference on empirical methods in natural language processing*. ACL, 1535–1545.
- [18] Katerina Frantzi, Sophia Ananiadou, and Hideki Mima. 2000. Automatic recognition of multi-word terms: the c-value/nc-value method. *International journal on digital libraries* 3, 2 (2000), 115–130.
- [19] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. 2017. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 1263–1272.
- [20] Ralph Grishman, David Westbrook, and Adam Meyers. 2005. NYU’s English ACE 2005 system description. *ACE* 5 (2005).
- [21] Zhou GuoDong, Su Jian, Zhang Jie, and Zhang Min. 2005. Exploring various knowledge in relation extraction. In *Proceedings of the 43rd annual meeting on association for computational linguistics*. ACL, 427–434.
- [22] Rahul Gupta, Alon Halevy, Xuezhi Wang, Steven Whang, and Fei Wu. 2014. Biperpedia: An ontology for search applications. (2014).
- [23] Keld Helsgaun. 2000. An effective implementation of the Lin-Kernighan traveling salesman heuristic. *European Journal of Operational Research* 126, 1 (2000), 106–130.
- [24] Ruihong Huang and Ellen Riloff. 2012. Bootstrapped training of event extraction classifiers. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 286–295.
- [25] Zhiheng Huang, Wei Xu, and Kai Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. *arXiv preprint arXiv:1508.01991* (2015).
- [26] Heng Ji and Ralph Grishman. 2008. Refining event extraction through cross-document inference. In *Proceedings of ACL-08: HLT*. 254–262.
- [27] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
- [28] Joseph A Konstan. 2008. Introduction to recommender systems. In *2008 ACM SIGMOD International Conference on Management of Data 2008, SIGMOD’08*. 1376776.
- [29] Terry Koo, Xavier Carreras, and Michael Collins. 2008. Simple semi-supervised dependency parsing. In *Proceedings of ACL-08: HLT*. 595–603.
- [30] Georgia Koutrika. 2018. Modern Recommender Systems: from Computing Matrices to Thinking with Neurons. In *Proceedings of the 2018 International Conference on Management of Data*. ACM, 1651–1654.
- [31] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360* (2016).
- [32] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. 2015. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.
- [33] Angli Liu, Stephen Soderland, Jonathan Bragg, Christopher H Lin, Xiao Ling, and Daniel S Weld. 2016. Effective crowd annotation for relation extraction. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 897–906.
- [34] Bang Liu, Weidong Guo, Di Niu, Chaoyue Wang, Shunnan Xu, Jinghong Lin, Kunfeng Lai, and Yu Xu. 2019. A User-Centered Concept Mining System for Query and Document Understanding at Tencent. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (Anchorage, AK, USA) (KDD ’19)*. ACM, New York, NY, USA, 1831–1841. <https://doi.org/10.1145/3292500.3330727>

- [35] Bang Liu, Di Niu, Kunfeng Lai, Linglong Kong, and Yu Xu. 2017. Growing story forest online from massive breaking news. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 777–785.
- [36] Jialu Liu, Jingbo Shang, Chi Wang, Xiang Ren, and Jiawei Han. 2015. Mining quality phrases from massive text corpora. In *SIGMOD*. ACM, 1729–1744.
- [37] Shulin Liu, Yubo Chen, Shizhu He, Kang Liu, and Jun Zhao. 2016. Leveraging framenet to improve automatic event detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2134–2143.
- [38] Xiao Liu, Zhunchen Luo, and Heyan Huang. 2018. Jointly multiple events extraction via attention-based graph information aggregation. *arXiv preprint arXiv:1809.09078* (2018).
- [39] Xueqing Liu, Yangqiu Song, Shixia Liu, and Haixun Wang. 2012. Automatic taxonomy construction from keywords. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1433–1441.
- [40] David McClosky, Mihai Surdeanu, and Christopher D Manning. 2011. Event extraction as dependency parsing. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Association for Computational Linguistics, 1626–1635.
- [41] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *EMNLP*.
- [42] Bhaskar Mitra, Fernando Diaz, and Nick Craswell. 2017. Learning to match using local and distributed representations of text for web search. In *WWW*. International World Wide Web Conferences Steering Committee, 1291–1299.
- [43] Makoto Miwa, Rune Sætre, Jin-Dong Kim, and Jun’ichi Tsujii. 2010. Event extraction with complex event classification using rich features. *Journal of bioinformatics and computational biology* 8, 01 (2010), 131–146.
- [44] David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Linguisticae Investigationes* 30, 1 (2007), 3–26.
- [45] Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 300–309.
- [46] Thien Huu Nguyen and Ralph Grishman. 2016. Modeling skip-grams for event detection with convolutional neural networks. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 886–891.
- [47] Aditya Parameswaran, Hector Garcia-Molina, and Anand Rajaraman. 2010. Towards the web of concepts: Extracting concepts from large datasets. *Proceedings of the VLDB Endowment* 3, 1-2 (2010), 566–577.
- [48] Youngja Park, Roy J Byrd, and Branimir K Boguraev. 2002. Automatic glossary extraction: beyond terminology identification. In *COLING 2002: The 19th International Conference on Computational Linguistics*.
- [49] Marius Pasca and Benjamin Van Durme. 2008. Weakly-supervised acquisition of open-domain classes and class attributes from web documents and query logs. In *Proceedings of ACL-08: HLT*. 19–27.
- [50] Sachin Pawar, Girish K Palshikar, and Pushpak Bhattacharyya. 2017. Relation Extraction: A Survey. *arXiv preprint arXiv:1712.05191* (2017).
- [51] Vasin Punyakanok and Dan Roth. 2001. The use of classifiers in sequential inference. In *Advances in Neural Information Processing Systems*. 995–1001.
- [52] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250* (2016).
- [53] Xiang Ren, Zeqiu Wu, Wenqi He, Meng Qu, Clare R Voss, Heng Ji, Tarek F Abdelzaher, and Jiawei Han. 2017. Cotype: Joint extraction of typed entities and relations with knowledge bases. In *WWW*. International World Wide Web Conferences Steering Committee, 1015–1024.
- [54] Alan Ritter, Sam Clark, Oren Etzioni, et al. 2011. Named entity recognition in tweets: an experimental study. In *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, 1524–1534.
- [55] Alan Ritter, Oren Etzioni, Sam Clark, et al. 2012. Open domain event extraction from twitter. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1104–1112.
- [56] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2017. Modeling Relational Data with Graph Convolutional Networks. *arXiv preprint arXiv:1703.06103* (2017).
- [57] Burr Settles. 2009. *Active learning literature survey*. Technical Report. University of Wisconsin-Madison Department of Computer Sciences.
- [58] Lei Sha, Feng Qian, Baobao Chang, and Zhifang Sui. 2018. Jointly extracting event triggers and arguments by dependency-bridge RNN and tensor-based argument interaction. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- [59] Jingbo Shang, Jialu Liu, Meng Jiang, Xiang Ren, Clare R Voss, and Jiawei Han. 2018. Automated phrase mining from massive text corpora. *IEEE Transactions on Knowledge and Data Engineering* 30, 10 (2018), 1825–1837.
- [60] Alisa Smirnova and Philippe Cudré-Mauroux. 2018. Relation Extraction Using Distant Supervision: A Survey. *ACM Computing Surveys (CSUR)* 51, 5 (2018), 106.
- [61] Yan Song, Shuming Shi, Jing Li, and Haisong Zhang. 2018. Directional skip-gram: explicitly distinguishing left and right context for word embeddings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. 175–180.
- [62] Frank Spitzer. 2013. *Principles of random walk*. Vol. 34. Springer Science & Business Media.
- [63] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 697–706.
- [64] Marco A Valenzuela-Escárcega, Gus Hahn-Powell, Mihai Surdeanu, and Thomas Hicks. 2015. A domain-independent rule-based framework for event extraction. In *Proceedings of ACL-IJCNLP 2015 System Demonstrations*. 127–132.
- [65] Kazufumi Watanabe, Masanao Ochi, Makoto Okabe, and Rikio Onai. 2011. Jasmine: a real-time local-event detection system based on geolocation information propagated to microblogs. In *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2541–2544.
- [66] Wikipedia contributors. 2019. Bipartite graph – Wikipedia, The Free Encyclopedia. [https://en.wikipedia.org/w/index.php?title=Bipartite\\_graph&oldid=915607649](https://en.wikipedia.org/w/index.php?title=Bipartite_graph&oldid=915607649). [Online; accessed 29-September-2019].
- [67] Ian H Witten and Olena Medelyan. 2006. Thesaurus based automatic keyphrase indexing. In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL’06)*. IEEE, 296–297.
- [68] Ian H Witten, Gordon W Paynter, Eibe Frank, Carl Gutwin, and Craig G Nevill-Manning. 2005. Kea: Practical automated keyphrase extraction. In *Design and Usability of Digital Libraries: Case Studies in the Asia Pacific*. IGI Global, 129–152.
- [69] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q Zhu. 2012. Probase: A probabilistic taxonomy for text understanding. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*. ACM, 481–492.

- [70] Bo Xu, Yong Xu, Jiaqing Liang, Chenhao Xie, Bin Liang, Wanyun Cui, and Yanghua Xiao. 2017. Cn-dbpedia: A never-ending chinese knowledge extraction system. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 428–438.
- [71] Yulan Yan, Naoaki Okazaki, Yutaka Matsuo, Zhenglu Yang, and Mitsuru Ishizuka. 2009. Unsupervised relation extraction by mining wikipedia texts using information from the web. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*. ACL, 1021–1029.
- [72] Bishan Yang and Tom Mitchell. 2016. Joint extraction of events and entities within a document context. *arXiv preprint arXiv:1609.03632* (2016).
- [73] Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piecewise convolutional neural networks. In *EMNLP*. 1753–1762.
- [74] Ziqi Zhang, José Iria, Christopher Brewster, and Fabio Ciravegna. 2008. A comparative evaluation of term recognition algorithms.. In *LREC*, Vol. 5.