S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

173

SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS INFORMATION NETWORKS

ISTRAŽIVANJE O POVEZIVANJU ENTITETA ZA SPECIFIČNE DOMENE S HETEROGENIM INFORMACIJSKIM MREŽAMA

S.Mythrei, S.Singaravelan

Department of Computer Science and Engineering, PSR Engineering College, Sivakasi, India
Odjel za računarske znanosti i inženjerstvo, PSR Engineering College, Sivakasi, Indija

Abstract

Entity linking is a task of extracting information that links the mentioned entity in a collection of text with their similar knowledge base as well as it is the task of allocating unique identity to various entities such as locations, individuals and companies. Knowledge base (KB) is used to optimize the information collection, organization and for retrieval of information. Heterogeneous information networks (HIN) comprises multiple-type interlinked objects with various types of relationship which are becomes increasingly most popular named bibliographic networks, social media networks as well including the typical relational database data. In HIN, there are various data objects are interconnected through various relations. The entity linkage determines the corresponding entities from unstructured web text, in the existing HIN. This work is the most important and it is the most challenge because of ambiguity and existing limited knowledge. Some HIN could be considered as a domain-specific KB. The current Entity Linking (EL) systems aimed towards corpora which contain heterogeneous as web information and it performs suboptimally on the domain specific corpora. The EL systems used one or more general or specific domains of linking such as DBpedia, Wikipedia, Freebase, IMDB, YAGO, Wordnet and MKB. This paper presents a survey on domain-specific entity linking with HIN. This survey describes with deep understanding of HIN,

Sažetak

Povezivanje entiteta je zadatak izvlačenja podataka koji povezuju spomenuti entitet u zbirci teksta sa njihovom sličnom bazom znanja, kao i zadatak dodjeljivanja jedinstvenog identiteta različitim entitetima, kao što su lokacije, pojedinci i tvrtke. Baza znanja (BZ) koristi se za optimizaciju prikupljanja, organizacije i pronalaženja informacija. Heterogene mreže informacija (HMI) obuhvaćaju višestruke međusobno povezane objekte različitih vrsta odnosa koji postaju sve popularniji i nazivaju se bibliografskim mrežama, mrežama društvenih medija, uključujući tipične podatke relacijske baze podataka. U HMI-u postoje razni podaci koji su međusobno povezani kroz različite odnose. Povezanost entiteta određuje odgovarajuće entitete iz nestrukturiranog teksta na webu u postojećem HMI-u. Ovaj je rad najvažniji i najveći izazov zbog nejasnoće i postojećeg ograničenog znanja. Neki se HMI mogu smatrati BZ-om specifičnim za domenu. Trenutni sustav povezivanja entiteta (PE) usmjeren je prema korpusima koji sadrže heterogene informacije kao web informacije i oni djeluju suptimalno na korpusima specifičnim za domenu. PE sustavi koristili su jednu ili više općih ili specifičnih domena povezivanja, kao što su DBpedia, Wikipedia, Freebase, IMDB, YAGO, Wordnet i MKB. U ovom radu predstavljeno je istraživanje o povezivanju entiteta specifičnog za domenu sa HMI-om. Ovo istraživanje opisuje s dubokim razumijevanjem

174

which includes datasets,types and examples with related concepts.

Introduction

Now-a-days, the amount of web data has been increased exponentially and it becomes one of the largest data repositories due to the increased amount of data. In the web, plenty amount of data forms the natural language. But, this language is highly uncertain based on named entities (NE) of frequent occurrences. The NE may contain various names and each name denotes various named entities. Most of the real systems usually comprised of muti-typed components which has a large number of interaction namely, social activities of human, computer systems, biological networks and as well as communication among networks. The INA (Information Network Analysis) has enhanced more attention in various specialists namely social, physics and engineering. Specifically, it has become the important most recent subject in data mining and retrieval field. The general approach is to retrieve hidden patterns through links from network of data.

EL determines the identity of the entity indicated in web text, and it is one of the issues in the bridging of unstructured data with the structured knowledge base and semi-structured knowledge base. The domain-specific KB is effective to maintain the query knowledge in a specific domain. Considering an example, IMDB contains more comprehensive knowledge about source of celebrity, movie and TV content than general KB named BaiduBaike and Wikipedia. The domain-specific Entity linking is the most important, in increasing demand for domain-specific knowledge base.

At the same time, some HIN would consider as domain-specific KB /**1**/. The DBLP network consist more information compared to Wikipedia based on the domain of computer science. The mapping of named entity with the DBLP knowledge base is meant as domain-specific EL. In this survey, we focused on entity linking on unstructured web-text in specific domain

HMI-a, što uključuje skupove podataka, vrste i primjere s povezanim konceptima.

that relates to the same domain as the HIN. This task automatically elucidates the domain-specific text with specific knowledge which is extracted from the HIN. The potential applications are question answering, information retrieval, information extraction, content analysis, and knowledge base analysis. However, these tasks are difficult due to entity ambiguity and name variations. This paper presents various survey of heterogeneous information network analysis, and domain-specific entity linking with the HIN.

The main objective of this survey is summarized as follows:

1.   To study the brief information of Heterogeneous Information Network (HIN) with various types of information networks.
2.   To analyze the Entity Linking (EL) and domain-specific entity linking.
3.   To study the survey of domain specific entity linking with various HIN.

The survey initiates an in-depth understanding of Heterogeneous information networks with various types of dataset information and techniques of HIN. It also investigates the various researches about meta path and meta structure in HIN. Section 3 contains surveys of Entity Linking and domain-specific entity linking. Section 4 describes the various surveys of Domain-Specific Entity Linking (DSEL) with HIN. Finally, section 5 contains conclusion of this paper.

Basic concepts of Heterogeneous Information Network

Heterogeneous Information Networks (HIN) defines the semantic network comprising various kinds of relations and entities /**2**/. The heterogeneous information network includes bibliographic networks, social networks, and do-

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS
INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

175

main-specific knowledge base. This section defines the important concept of HIN and techniques of HIN in various literature surveys.

A) Information network

Generally, an information network defines real world abstraction, focuses on the types of relationships between the objects. It is not only about storing and representing important real world information, but also provides a efficient tool to retrieve the knowledge from it. An information network is ubiquitous and prevalent in the way of modeling/representing in interacting components.

The Information Network (IN) is represented as a projected graph in the form of $G = (V, E)$ and mapping function: $V \rightarrow$ A, which denotes object mapping where an object related to only one type and E$\rightarrow$ R, denotes link mapping where an link belongs to one relation type i.e each link explicates the specific relation. When two links contain the same relation type, these two links divides the same starting object type and same ending object type. When the object type $|A| > 1$ and relation type $|R| > 1$, the network is called **Heterogeneous information network**, otherwise, it is a **Homogeneous information network**.

B) Network schema

The network schema is represented as $T_G =$ (A, R) and it defines the Meta template for heterogeneous network $G = (V, E)$ with mapping function φ.

Here, A $\rightarrow$ object type and R $\rightarrow$ Edges relation. For directed graph V represents the mapping of object type A. And E represents the edges relation from R.
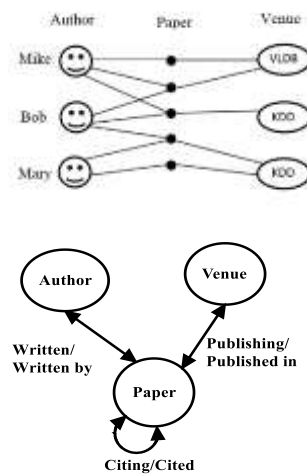

Figure 1: Example for Network instance and Network schema

Figure 1 represents the instances of HIN for bibliographic network and bibliographic network schema. In network, the network schema is followed by network instances. The heterogeneous information network /**3**/ is arranged by multiple interconnected and huge datasets that range from scientific, social, engineering and business applications. Some of the networks are

1. Bibliographic information network
2. Twitter information network
3. Flickr information network
4. Healthcare information network
1. **Bibliographic information network**

In bibliographic networks, the multiple types of an object contain *authors (A), Papers (P), Publication venues (V)* and *title terms (T)* which are interconnected together and provides rich information named Heterogeneous information networks /**4**/. But, the object names are potentially ambiguous in heterogeneous information network which means the same named textual refers to various several entities. For example, in the network of DBLP, the network "Wei wang" refers to fifty various authors that include various universities' names, city, state and other required information. However, there are large existing heterogeneous information networks contains less information. In DBLP network, the "advisor" relation does not exist between authors /**2**/. Therefore, the most popular of exist-

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS
INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

176

ing HIN with newly extracted information's become most important. In general, entities of the information are comprehensively interlinked with one another through a different field.
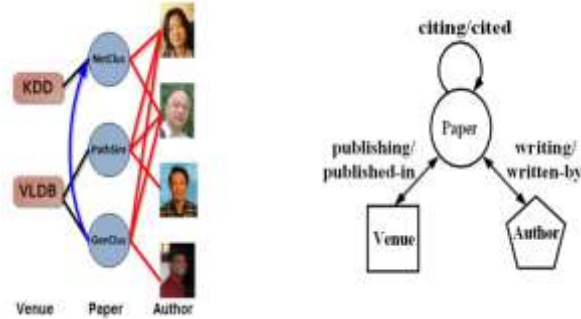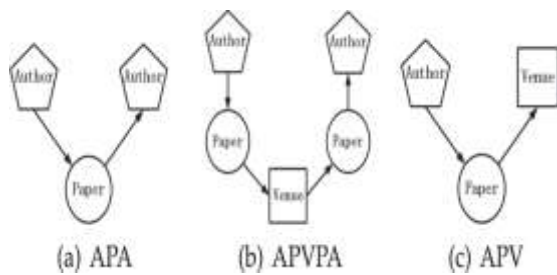

Figure 2: Example for HIN on bibliographic dat /2/

Figure 2 depicts the HIN representation on bibliographic dataset.

C) Meta path

A Meta path α is defined as the path represents on a Heterogeneous information network schema. In the network schema S = (A, R), Meta path is denoted as $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} A_3 \xrightarrow{R_3} A_{l+1}$. Here R_1, R_2 and R_3 represents the composite relationship between the objects of A_1 and A_{l+1}. Each Meta path has different meanings. The number of relations in α is the length of the meta path α.

Figure 3: Example for Meta path in HIN on the bibliographic network /2/

Figure 3 defines the meta path representation in HIN on bibliographic network. In this figure APA refers "Author-Paper-Author. APVPA refers Author-Paper-Venue-Paper-Author" and APV refers Author-Paper-Venue. Meta path connects one or more objects. The APA path defines the author's collaboration in same paper.

The APVPA defines the author publishing papers on same venue. In APV path, the author and venue connected in same paper.

2. Twitter information network

This type of network is also called a social media information network that contains the objects of user, hash tag, and tweet. The link types are defined as *follow*ing users, *post*, *reply*, and *use*.

3. Flickr information network

Flickr information network is a photo-sharing information network, and it object types of it's are: *user, image, tag, group,* and *comment*. The link types are defined as *upload images; belong to, post,* and *comment.*

4. Healthcare information network

The healthcare information's network defined as the objects types of *doctor, patient, disease, and treatment*. And the link types are referred to as *used for, visit* and *device.*

D) Techniques of Heterogeneous Information Networks (HIN)

The heterogeneous information network introduces the mining task which includes the tasks of ranking, clustering, classification, similarity search, recommendation, relationship search, and link prediction.

/5/ described the HIN embedding for the recommendation in both exploitation and extraction. This research proposed a Novel Heterogeneous network embedding based method for Recommendation which is known as HERec. It defines the HIN was agreed to categorize the auxiliary and complex data in recommender systems because of the flexibility in data heterogeneity, called HIN based recommendation. For network embedding, it provides the strategy of random walk off based meta-path was designed to estimate the node sequences. For both extraction and exploitation, the effective method for HIN based recommendation was developed. By the group of fusion functions, the embedding nodes are transformed and consequently, integrated into an enlarged matrix factorization model. Finally, the enlarged matrix factorization was connectedly optimized for rating

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS
INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

177

prediction performance. In comparison a-nalysis, the proposed method was compared with existing HIN based recommendation methods. For experimental results, there are three types of datasets are demonstrated for effective output. Also the proposed HERec method mitigates the cold start issues and examines the meta-paths impacts on perfor-mance. The disadvantage of this paper is, it doesn't investigate the deep learning methods in multiple meta-paths.

/6/ Demonstrated the similarity search in Schema-Rich HIN. There were more studies explained about real-world data as HIN conta-ins multiple types of relations and objects. But some of the studies restrict the network discus-sions with entities and relationship types. The network schema was more complex in the real world such as freebase. In some times, HIN in the rich schema can contain more complex and more burdens to provide information to users. Under some setting issues, users are asked to give some simple instance relation samples as a query and predict the latent semantic relation (LSR) inferred by the query which aided to iden-tify other relation instances. To overcome this problem a new meta-path based relation and Relsim were defined to analyse the similarity between the instances. This paper proposed two requirements based on meta-path and Realism. Firstly, they represented an LSR with query-ba-sed path that generates based on the network schema. Secondly, a novel meta-path based re-lation with Relsim was introduced and utilized the inefficient search similarity algorithm.

/7/ Provided the framework for dynamic LP (link prediction) in HIN. This study proposed an effective two-level scheme that produces competent macro-decisions and micro-decisi-ons for combine structure, time-sensitive and dynamic content. It proposed a dynamic link in-ference algorithm in temporal and heterogene-ous networks. The proposed algorithm was a-ble to construct the link inference model for he-terogeneous networks. The main advantage of this research is to describe the content-based model and structure-based model based on a

dynamic clustering approach. The time-sensi-tive nature is to evaluate effective link predic-tion. Network and linked data were quite pre-valent because of the popularity of the web and social media applications which are essentially network-oriented. The main purpose was to take a unique technology such as large scale dynamic link prediction for higher-level decisi-ons using topological behavior and also for more fine-grained decisions. It also studied the problem of community prediction. Finally, explained the effectiveness of the proposed technique over several real datasets.

/8/ explained the entity recommendation in the HIN. This paper studied the HIN recommenda-tions and proposed a generic framework for the implicit dataset by utilizing the different types of entity relationships. Firstly, it introduced Meta path-based features to estimate the connection between the users and items in various paths. This paper described recommendation models at both personalized and global levels. The perso-nalized model was effectively produced on the fly and provided high-quality results compared to other systems. This study utilized a Bayesian ranking process to evaluate the weights of re-commendation models. This empirical study de-monstrated the effectiveness of the method and examined the performance of this method under different consequences. Furthermore, the study will explain the on-line recommendation model update with large scale recommendation and le-arning process along with low time complexity. In comparison analysis, the proposed method was compared with various state-of-the-art met-hods and various feedback recommendation methods. The evaluation results are examined in the IMDb dataset for effective performances.

/9/ Provided the Meta structure for computing si-milarity or relevance in a large set of HIN. This research utilized Meta structure which is a focu-sed acyclic graph to evaluate the proximity between objects with edge types connecting in between. This research introduced the notion of meta-structure, and it defines the extension of the Meta path. They defined two relevant mea-sures such as SCSE and StructCount. SCSE tri-ggers the sub graph expansion and reduces the

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS
INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

178

bias to visible objects. They defined BSCSE which combines both StructCount and SCSE. To improve BSCSE efficiency, this research proposed a recursive algorithm based on two optimizations such as i-LTable and Compressed-ETree. The main key of the Meta structure described a complex relationship between two HIN objects. This approach modeled large and complex datasets namely YAGO and DBLP as HINs. The main issues are closeness between two objects. Investigated studies based on HIN information utilized for significance computation. Based on Meta structure three significance measures are measured. Because of the occurrences of the complexity, these metrics further design an algorithm with the data structure in a proper format which gives more efficiency than another existing approach.

/**10**/ explicated the meta-path search and data mining in HIN. This study also discussed issues arose from the search and mining of HIN.The Meta path is a powerful tool to define the relationship between two objects. This research proposed a Meta path concept to capture the semantic relationship between various types of objects. For mining the heterogeneous information network this study proposed Meta path-based search. This Meta path guides mining of the network and searching network, it also understands the semantic meaning of network objects. Different functions are proposed and developed namely community detection, link prediction, and ranking. It addressed the relationship prediction as well as user-guided clustering. Using this meta-structure or path this approach generates high accuracy and provided better communication toward the relations between the user-specific tasks and finally introduced several research frontiers. The HIN network, it described as a bibliographic information network, twitter information network, flicker information network, and healthcare information network. It also explains the data mining applications in network meta-structure, and the applications are clustering, similarity search, recommendation, and link prediction.

/**11**/ Proposed the method for patent trading which was enhanced by HIN. The Patent information was defined as patent citation, patent trading, patent contents and patent invention are integrated with HIN. This proposed technique was examined with online and offline experiments. There are some limitations which can be improved further. This system identifies the IP interests of companies depend on the online behaviors.

/**12**/ Suggested heterogeneous information network creation using classical similarity measure, attributes of document and terminology products with the help of Network creator algorithm. The similarity measures of documents in HIN were measured by HeteSimTKS Query. The results show that this algorithm is feasible but the runtime of network construction takes much time.

Domain-specific entity linking

This section explains the various researches and studies about Entity Linking (EL) in specific domains and various issues in EL.
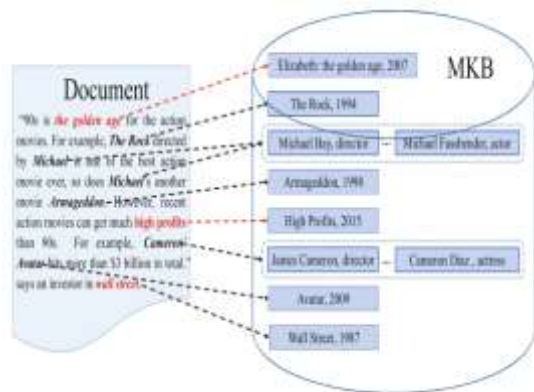
A) Entity linking

Entity linking mainly aims to map the textual mentions with corresponding entity text in a knowledge base by solving the problem of semantic ambiguity. It determines the identity of the entity in text, and it is the key issue in the unstructured data bridging with the structured knowledge base and semi-structured knowledge base /**13**/. The domain-specific KB contains more effective to maintains the query knowledge in relevant domain.

B) Domain-specific Entity Linking

The EL systems used one or more open domains for linking such as DBpedia, Wikipedia, Freebase, IMDB, YAGO, DBLP,Craft,Genia,Wordnet and MKB, based on their disambiguation models and entity recognition. The training of disambiguation models and entity recognition on the corpus could contain high costs. Instead of this, the entity linking can be used on a corpus for training without modification. And the performance depends on the similarity content of the corpus (EX, Wikipedia). The current Entity Linking systems are directed over corpora that are defined as hete-

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS
INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

179

rogeneous as web /**14**/. So it provides sub-optimal results on domain-specific corpora. The applications of entity linking defined as,

1. Information retrieval
2. Information extraction
3. Content analysis
4. Question answering

Figure 4: Representation of domain-specific entity linking /15/

In this figure 4 defines with Movie-Knowledge Base(MKB) and the document contents directly linked into MKB database. In document, the common phrases are "the golden age", "high profits" and "wall street". "The rock" which is an actual named entities in domain specific MKB database.

/**16**/ explained the adaptable entity linking, a hybrid method to link the entities for information extraction. This research proposed a method that allows indexing the data independently and vocabulary was being used. The main purpose was to design the index so that each entity had two information namely popularity score and page rank score. Four main issues cause numerous difficulties when establishing an EL system such as the number of types used to categories an entity, kind of textual documents, the knowledge base used to remove the extracted data and the language utilized in the documents. Finally, the framework was analyzed named entity recognition, ADEL, and linking system which is based on a hybrid lin-

guistic semantic-based and information retrieval. This framework was evaluated using six datasets such as OKE2016, OKE2015, NEEL2015, and NEEL2016 AND AIDA. This showed that the indexing approach enables to produce an accurate group of resources from any KB in minimum time with a minimal size.

/**17**/ Proposed Named Entity Linking (NEL) methods focused on techniques particularly developed to handle tiny messages namely tweets. NEL is the mission of the annotation of entity mentioned in a text portion with links to a KB and utilized for content extraction and tweets annotation that are essential for making large amount of users in Twitter's who are interconnected. The range of one hundred and forty characters associates a very tiny and noisy context that causes new issues to the EL on twitter. Since the text length could not exceed 140 characters. The texts are tough to remove the uncertainty even for human readers. Finally, this approach provided a framework for the performance of NEEL during # micro-posts NEEL issues from previous decades. Additional data are identified in the dataset established by Microsoft. Furthermore, this paper haven't discussed about the new metrics design at different linguistic level.

/**18**/ Introduced Entity linking in two different domains with different objectives that involved article wide and article wide –salient and instance-based EL and challenges of each type of entity linking. There are various researches showed the specifying information about the certain entity and it requires to analyze the properties of specific entities. This paper compares the general entity linking system architecture and general normalization system architecture. This EL which was apart from the entity recognition tasks using linking textual entity to KB entry. This paper surveyed in general as well as the biomedical domain. Furthermore, the result of the EL task is given for references that uncover the new challenges identified in the mining of biomedical text.

/**19**/ presented an extensive computation based on nine datasets and proposed an unsupervised system named VINCULUM for entity linking.

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS
INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

180

To rectify the confusion of the entity linking problem, this paper presented with the VINCULUM method. Even though numerous papers are described in EL with several problems. In this paper, nine EL datasets are examined to improve a better understanding of an EL and to measure the impact of each component. The evaluation results are examined in nine types of datasets. This showed strong candidate generation components which provided good results using fine-grained entity types to filter out the wrong links. Finally, this system achieved comparable performance with existing machine learning linking systems. It analyzes the difference between various entity linking problems and compares the various existing datasets. They investigated several key aspects of the system such as entity type prediction, mention extraction, entity coreference, coherence between entities and candidate generation.

/20/ Proposed a probabilistic model which determined user-generated information based user on queries to KB entities. To introduce a contextual relevance model, uses learned query words and entities. The relative measure is quickly computed among an entity and the string of text. EL is a process with the identification of entities in a given piece of text for better web document ranking. In this model, three key components made the model fast and efficient. First, the dependencies between the different entities are avoided during the linking process which enabled the implementation in query terms. Second, to decrease the memory footprint that influenced the compression and hashing techniques. Finally to prepare algorithm with contextual knowledge and factored the distance between the query cords and entities. The main problem in this system is that the process wanted to run below the time constraints which were performed before any actual retrieval.

Domain-specific entity linking with hin

In HIN, there are multi-connected networks and interconnected objects. The linking named entity task involved in the heterogeneous information network. In HIN, let us consider the knowledge base ex: Freebase. In this knowledge

base entity, there are many entities are organized in hierarchy categories. Each entity contains additional information in large HIN. /21/ Described the knowledge base entity linking with techniques, issues, and solutions. The various applications are presents for bridging web data with the knowledge base was increased in the EL research. They utilized three factors of entity linking systems such as generation of candidate entity, ranking of candidate entity, and unlinkable mention prediction. This research presented with many efforts in entity linking. It is a task to link entities with their relevant entities. Potential applications involve content extraction, information retrieval features, application, and evaluation. Finally, it presented many determinations in entity linking.

/22/ Proposed the probabilistic model for HIN entity linking with web text. The named entity linking task was determined from unstructured data with their related HIN entities. And this task was the most challenge due to limited knowledge in existing HIN. To recover this issue, this paper proposed a SHINE model to connect the web text named entities with HIN. This model comprises two components entity popularity and object model. This model captured the entity popularity and the handling of multi-type objects model appeared in the textual context. The multi-type objects distribution was estimated using multi-path constrained random walks through networks. Various meta paths expressed as a different distribution over networks. An effective iterative method to study for every meta-path weights which is based on the Expectation-Maximization(EM) algorithm. This paper clearly explained DBLP and IMDb network schema in HIN and the probabilistic model of EL with HIN.

/23/ Proposed the Literature-based discovery. A unique Literature-Based Discovery (LBD) named HBIN – LBD has been designed the influences lexical citation structures in HBIN graphs. The effects of implementing topic and semantic modeling into the proposed technique have been discussed. Proposed LBD method determines the meta-path features in HBIN graphs for latent association's determination unlike

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

181

existing LBD methods. A drawback of this method is that it doesn't explore the efficient graph algorithm against large HBIN graphs.

/**24**/ This study proposed knowledge-base-agnostic approach. It describes a knowledge base method for named entity rendition. It detects the correct URI within an input text and used eight different datasets for a group of named entities. It Profits from developing KN as well as multilingual linked data. An advantage of this model is that proposed AGDISTIS can effectively determine the accurate URIs for a group of named entities. And it doesn't extend with more domain-specific datasets.

/**25**/ Proposed the named entity linking. It described the compact context representation which contains dependency features for improving NEL quality. Given the extracted mentions from the document entity linking systems and identify the set of candidates from Wikipedia. It identified the different designs of context representation. This method affiliates the high-level dependency in similarity measures. For these informative features, it exists a lack of overlap between a mentioned-entity.

/**26**/ Proposed a knowledge-base-agnostic approach. It contains the global EL approach and performed the page rank algorithm which is to perform a random walk with the restart. It Measured coherence among the entities and documented them in a unified manner. The direct connected entities assumed as the semantic representation of entities, the semantic entities are improves the results of entity linking and provides better performance compared to other methods.

/**27**/ Evaluated the Roadmaps of an active research group. Describes the link prediction techniques that identify the missing links in recent networks used. Topology and learning-based methods are used to find the missing links. Social network data are incomplete and it is highly dynamic which cause the nodes and edges in the graph to visible or invisible for upcoming applications. Link prediction determines the missing link inefficient way in network.

/**28**/ Explicated the Meta path similarity. Estimated the link using meta path mutual information index. Instead of node-pairs connectivity, it evaluated the amount of information through paths. Proposed MMI method produces the maximum prediction accuracy when compared with existing methods. They do not consider inside the meta path when the same indices are strongly dependent the reachability amount.

/**29**/ Proposed Weighted heterogeneous information network. Described a Weighted HIN along with semantic linking (WHIN – CSL) has been devised for learning high-quality feature representations. The semantic link is computed using abstract similarity. The proposed method has high flexible integrating information, so it achieves the highest performance. Using the full text of papers, instead of using only abstracts, for computing semantic similarity for developing semantic linking can be commendable.

/**30**/ Proposed the Latent Dirichlet Allocation method. Hierarchical Expert profiling has been developed where the knowledge areas of the research scholar are plotted in various granularity levels. A modeling strategy has been designed for evaluating the Profile experts. The expert profiling problems and topic modeling problems are addressed. In the expert profiling domain, usage of node ranking can be taken into consideration.

/**31**/ discussed the data-driven natural language processing and machine learning. It incorporates the domain knowledge in social media network. This generates Bigdata marked by volume, velocity and variety challenges. This type of Bigdata analytics support useful investigation from data mining. Solving the problems and challenges of domain-specific linking in social media was the most required, and it occurs due to complexity of the domain, converse nature of language and change of topic relevance. It also integrates contextual interpretation of decision makers and experts. The challenges of this study are addressed using four use cases which are used to show the domain knowledge role.

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS
INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

182

/**32**/ Presented with Probabilistic entity-linking method. Describes a probabilistic algorithm has been designed by using knowledge graphs for solving entity linking problem. Most of the existing work only focus on the search log and the entity linking database and there may be complex graphs but the proposed algorithm works as simple, fast and accurate. The problem associated with entity linking is addressed in this study and provide link graph structure. The performance can be improved by textual evidence.

/**33**/ This study Generalized coupled tensor factorization. The link prediction problems are solved by using coupling analysis in the form of high order and tensors. A GCTF framework has been used for developing coupled models. It improved the performance via joint analysis, and efficient detection of missing links between entities. Estimation of dispersion parameters could be done in future and loss function had not been described, it consider as the disadvantage.

/**34**/ proposed the framework for domain specific EL with HIN. The existing EL methods only focused entity linking in Wikipedia. This paper presented SHINE+ method, and it defines the entitie"S" in HINEtwork. The proposed probabilistic model described with entity popularity model and entity object model along with knowledge population algorithm, if the network information was insufficient. This algorithm was utilized with high confidence to enhance the knowledge of entity by using context information which are mapped by linking model. It also boost the performance of linking. The results of two networks such as IMDb and DBLP show the proposed method framework and efficiency of the method. For solving the optimization problem, the gradient descent method was used. The SHINE+ approach had quite stable, high and insensitive over three datasets. Also the sensitivity analysis was performed to understand the proposed method parameters. Learning algorithm had time complexity when presents the enormous number of entities and it becomes little expensive.

/**35**/ explained the "THINKER" model for Entity Linking. The existing systems didn't explain about the entity linking in Turkish language. This paper proposed the novel entity linking for Turkish content linking as the name of THINKER which means the Turkish content and entities specified in Turkish dictionary such as r.wikipedia.org and tdk.gov.tr. The proposed system contains the challenges of Entity Disambiguation and Entity Detection. To address these challenges initially, it proposed machine learning based entity detection algorithm. Secondly, it presented with collective disambiguation algorithm for task of linking and optimization was done with genetic algorithm. It also described the link similarity and entity ranking estimation. The proposed system contains the challenges of Entity Disambiguation and Entity Detection. In result analysis the proposed method was estimated with two various spotting algorithms named THINKER spotter and N-gram spotter. For comparative analysis the proposed framework was compared with WikiME Systems. In this method, the correct entity mention was considered as undetected entity when it does not exist in knowledgebase.

Conclusion

The EL is an emerging area and gives challenges in several research work fields. Most of the interest domains are related towards linking entity as well as interrelation among multi type heterogeneous objects. For better understanding, We have surveyed the clear concepts of HIN, some datasets in HIN, Meta path or Meta structure in HIN and Entity linking with HIN along with the inspect of domain-specific EL with HIN. There are many researches proposed entity linking with HIN for general knowledge base instead of useful domain-specific knowledge bases. This survey is expected to be the most useful for researches to analyze in this field.

Notes

/1/ C. Wang, X. He, and A. Zhou, "HEEL: exploratory entity linking for heterogeneous information networks," *Knowledge and Information Systems*, pp. 1-22, 2019.

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS
INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

183

/2/ Y. Sun and J. Han, "Mining heterogeneous information networks: a structural analysis approach," *Acm Sigkdd Explorations Newsletter,* vol. 14, pp. 20-28, 2013.

/3/ F. Polzin, H. Toxopeus, and E. Stam, "The wisdom of the crowd in funding: information heterogeneity and social networks of crowdfunders," *Small Business Economics,* vol. 50, pp. 251-273, 2018.

/4/ C. Shi, Y. Li, J. Zhang, Y. Sun, and S. Y. Philip, "A survey of heterogeneous information network analysis," *IEEE Transactions on Knowledge and Data Engineering,* vol. 29, pp. 17-37, 2016.

/5/ C. Shi, B. Hu, W. X. Zhao, and S. Y. Philip, "Heterogeneous information network embedding for recommendation," *IEEE Transactions on Knowledge and Data Engineering,* vol. 31, pp. 357-370, 2018.

/6/ C. Wang, Y. Sun, Y. Song, J. Han, Y. Song, L. Wang, *et al.,* "Relsim: relation similarity search in schema-rich heterogeneous information networks," in *Proceedings of the 2016 SIAM International Conference on Data Mining,* 2016, pp. 621-629.

/7/ C. C. Aggarwal, Y. Xie, and P. S. Yu, "A framework for dynamic link prediction in heterogeneous networks," *Statistical Analysis and Data Mining: The ASA Data Science Journal,* vol. 7, pp. 14-33, 2014.

/8/ X. Yu, X. Ren, Y. Sun, Q. Gu, B. Sturt, U. Khandelwal, *et al.,* "Personalized entity recommendation: A heterogeneous information network approach," in *Proceedings of the 7th ACM international conference on Web search and data mining,* 2014, pp. 283-292.

/9/ Z. Huang, Y. Zheng, R. Cheng, Y. Sun, N. Mamoulis, and X. Li, "Meta structure: Computing relevance in large heterogeneous information networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 2016, pp. 1595-1604.

/10/ Y. Sun and J. Han, "Meta-path-based search and mining in heterogeneous information networks," *Tsinghua Science and Technology,* vol. 18, pp. 329-338, 2013.

/11/ Q. Wang, W. Du, J. Ma, and X. Liao, "Recommendation Mechanism for Patent Trading Empowered by Heterogeneous Information Networks," *International Journal of Electronic Commerce,* vol. 23, pp. 147-178, 2019.

/12/ A. A. P. Ribeiro, "Similarity algorithms for Heterogeneous Information Networks," Universidade de São Paulo, 2019.

/13/ A. Moro, A. Raganato, and R. Navigli, "Entity linking meets word sense disambiguation: a unified approach," *Transactions of the Association for Computational Linguistics,* vol. 2, pp. 231-244, 2014.

/14/ F. Hasibi, K. Balog, and S. E. Bratsberg, "Exploiting entity linking in queries for entity retrieval," in *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval,* 2016, pp. 209-218.

/15/ J. Zhang, J. Li, X.-L. Li, Y. Shi, J. Li, and Z. Wang, "Domain-specific entity linking via fake named entity detection," in *International Conference on Database Systems for Advanced Applications,* 2016, pp. 101-116.

/16/ J. Plu, G. Rizzo, and R. Troncy, "ADEL: ADaptable Entity Linking," *Semantic Web Journal,* 2019.

/17/ X. Liu, Y. Li, H. Wu, M. Zhou, F. Wei, and Y. Lu, "Entity linking for tweets," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers),* 2013, pp. 1304-1311.

/18/ H.-J. Dai, C.-Y. Wu, R. T.-H. Tsai, and W.-L. Hsu, "From entity recognition to entity linking: a survey of advanced entity linking techniques," in *26th Annual Conference of the Japanese Society for Artificial Intelligence (2012),* 2012, pp. 110-119.

/19/ X. Ling, S. Singh, and D. S. Weld, "Design challenges for entity linking," *Transactions of the Association for Computational Linguistics,* vol. 3, pp. 315-328, 2015.

/20/ R. Blanco, G. Ottaviano, and E. Meij, "Fast and space-efficient entity linking for queries," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining,* 2015, pp. 179-188.

/21/ W. Shen, J. Wang, and J. Han, "Entity linking with a knowledge base: Issues, techniques, and solutions," *IEEE Transactions on Knowledge and Data Engineering,* vol. 27, pp. 443-460, 2014.

/22/ W. Shen, J. Han, and J. Wang, "A probabilistic model for linking named entities in web text with heterogeneous information networks," in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data,* 2014, pp. 1199-1210.

/23/ Y. Sebastian, E.-G. Siew, and S. O. Orimaye, "Learning the heterogeneous bibliographic information network for literature-based discovery," *Knowledge-Based Systems,* vol. 115, pp. 66-79, 2017.

/24/ R. Usbeck, A.-C. N. Ngomo, M. Röder, D. Gerber, S. A. Coelho, S. Auer, *et al.,* "AGDISTIS-graph-based disambiguation of named entities

S.Mythrei, S.Singaravelan: SURVEY ON ENTITY LINKING FOR DOMAIN SPECIFIC WITH HETEROGENEOUS
INFORMATION NETWORKS
Informatologia, 52, 2019, 3-4

184

using linked data," in *International semantic web conference*, 2014, pp. 457-471.

/25/ X. Ling, S. Singh, and D. S. Weld, "Context representation for named entity linking," in *Pacific Northwest Regional NLP Workshop*, 2014.

/26/ Z. Guo and D. Barbosa, "Entity linking with a unified semantic representation," in *Proceedings of the 23rd International Conference on World Wide Web*, 2014, pp. 1305-1310.

/27/ P. Wang, B. Xu, Y. Wu, and X. Zhou, "Link prediction in social networks: the state-of-the-art," *Science China Information Sciences*, vol. 58, pp. 1-38, 2015.

/28/ H. Shakibian and N. M. Charkari, "Mutual information model for link prediction in heterogeneous complex networks," *Scientific reports*, vol. 7, p. 44981, 2017.

/29/ J. Chen, Y. Liu, S. Zhao, and Y. Zhang, "Citation Recommendation Based on Weighted Heterogeneous Information Network Containing Semantic Linking," in *2019 IEEE International Conference on Multimedia and Expo (ICME)*, 2019, pp. 31-36.

/30/ J. Silva, P. Ribeiro, and F. Silva, "Hierarchical Expert Profiling Using Heterogeneous Information Networks," in *International Conference on Discovery Science*, 2018, pp. 344-360.

/31/ S. J. Kho, S. Padhee, G. Bajaj, K. Thirunarayan, and A. Sheth, "Domain-specific Use Cases for Knowledge-enabled Social Media Analysis," in *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*, ed: Springer, 2019, pp. 233-246.

/32/ S. Bhatia, "Entity Linking in Enterprise Search: Combining Textual and Structural Information," in *Linking and Mining Heterogeneous and Multi-view Data*, ed: Springer, 2019, pp. 183-199.

/33/ B. Ermiş, E. Acar, and A. T. Cemgil, "Link prediction in heterogeneous data via generalized coupled tensor factorization," *Data Mining and Knowledge Discovery*, vol. 29, pp. 203-236, 2015.

/34/ W. Shen, J. Han, J. Wang, X. Yuan, and Z. Yang, "Shine+: A general framework for domain-specific entity linking with heterogeneous information networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, pp. 353-366, 2017.

/35/ M. Kalender and E. E. Korkmaz, "THINKER-Entity Linking System for Turkish Language," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, pp. 367-380, 2017.