IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 40-46

A Survey on Linking Locations Using User Comments

N.Mahesh[1], Dr.N.Muthumani[2]

Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, India[1]

Professor and Head, Department of Mathematics (CA), Sri Ramakrishna College of Arts and Science, Coimbatore, India[2]

Abstract: *Many websites are based on domain specific contents for profile hosting, for example: locations on foursquare, Products in flipkart, amazon and movies in IMDB using the comments post by users on. When user is comment on the particular entity the reference and comparison is made from each and every entities in it. Thus we often mention on entities with comparision makes an object and entity relation ship model that compared to other web pages, in tweets it disambiguate entities of comment by users that is yet to received, many of users reveal their focal linking locations such as land marks, shops, restaurants and mall in tweets, the problem arise in disambiguate with already mention entities of user comments which has not yet received. Fine grained locations make linking of entities easily of location profiles with the gio-coordinates. It is very often to define location profiles like geo co-ordinates, In existing system location linking and recognition has two methods such as pipeline setting, novel joint framework performance location recognition and linking of location simultaneously in state space search. The location linking problem gives an prediction problem arise of beam search based algorithm using concept learning of further learning using unlabeled data, using labeled data that the extensive research are done to recognize locations mentioned in tweets to location profile in search. The unlabeled data improve both linking and recognition accuracy for more good solutions.*

I. Introduction

Sentiment Analysis requires language resources such as dictionaries that provide a sentiment or emotion value to each word. Just as words have different meanings in different domains, the associated sentiment or emotion also varies. Hence, every domain has its own dictionary. The information about what each domain represents or how the entries for each domain are related is usually undocumented or implied by the name of each dictionary. Moreover, it is common that dictionaries from different providers use different representation formats. Thus, usage of different dictionaries is very complex at the parallel time.

Nowadays the many people need to coordinate this kind of innovation into their applications. What's more, with the ascent of AI, cloud computing and large scale information handling tools; it is never again reasonable for the tech monsters. That is the reason business answers for element connecting APIs have developed as of late.

Be that as it may, they particularly targeted on the 'named entity' extraction and miss different ideas. Named substances are ordinarily comprehended as locations, products or organizations, people.

Let’s see our examples to enhance recognization about the limitations. Existing APIs would just concentrate 'Yann LeCun' and 'CNN' notices. They would miss 'Deep learning', 'Artificial neural system'. They likewise would not disambiguate 'CNN' to 'Convolution neural system' since it's anything but a named substance and they could even connection it to the TV channel idea.

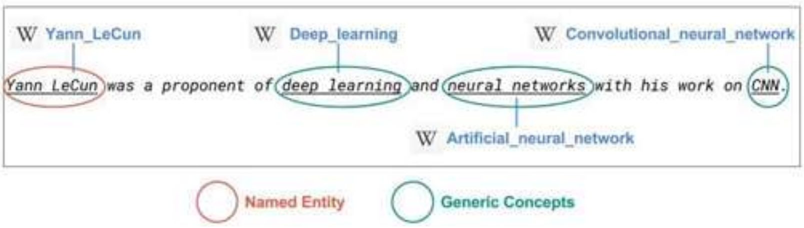

Figure 1.1. Named entity are only a fraction of concepts that can be extracted from text

These business APIs are going the correct way as they as of now enable individuals to fabricate better applications. For instance, that is an impeccable instrument in the event that you need to construct an application that screen what individuals are saying in regards to a lawmaker or a particular brand. But they can still be limited for more advanced text analysis and for more ambitious applications.

For instance, at asgard.ai, we fabricated our own extended linking system framework. Our fundamental objective is to enable individuals to all the more likely discover and track technologies and their ecosystems. We are improving this experience by organizing data through most recent achievement in natural language processing , intuitive interfaces and machine learning. With this objective in mind, we are building an AI-powered knowledge graph about technologies and their ecosystem by analyzing a lot of unstructured data (scientific publications, patents, company description, …). When we consequently dissect content, we need to extricate every single basic idea and discover the connections between them with a specific end goal to separate learning from it. That is the reason we extended substance connecting to all ideas and not simply 'named entities'.

1.1. Why extended substance connecting is urgent for important content examination

Human dialect isn't correct. Specifically, in the event that you endeavor to investigate and comprehend messages naturally, you'll confront a noteworthy test : how to deal with ambiguity and variability.

1.1.1. Variability

People don't generally utilize a similar word to refer to a similar idea or element. However it could be helpful to get archives in a corpus that allude to a particular hidden idea, whatever the manner in which people allude to them.

1.1.2. Ambiguity

Regularly peoples are utilizing questionable notices to allude to ideas or elements. It is on the grounds that the encompassing setting of the word or expression contains enough data for the reader to disambiguate and comprehend fundamental ideas. In phonetics, these concealed data are called logical signals.

Here have two classic approach for keywords and extraction of phrases, that extract and handling of issues directly, with deal of statistical modeling it is also called as distributional semantic models. The word in context or documents will handle for document based similarities, they have two approaches namely 1. context counting model and 2. context predicting models they are explained below:

Context counting model

In this methods word 'count' is an co-occurrence were documents as in corpus. That gives an model based latent topics and distribution of word to likely share and co-occurance based sharing of advanced topic modeling, it is also have to compute "vector based representation of words" also called as word embedding method.

Context predicting models

In here the method predict a word of context for the purpose of word embedding.

Above mentioned approach gives an work well of result ranking for document computation similarities to achieve an high level classification approach. But it is not only sufficient to use those cases of doing statistics of specific data set and extraction of knowledge from texts. It is the reason of "adding an entity linking component" to this approaches is very essential and ambiguity underlying concepts.

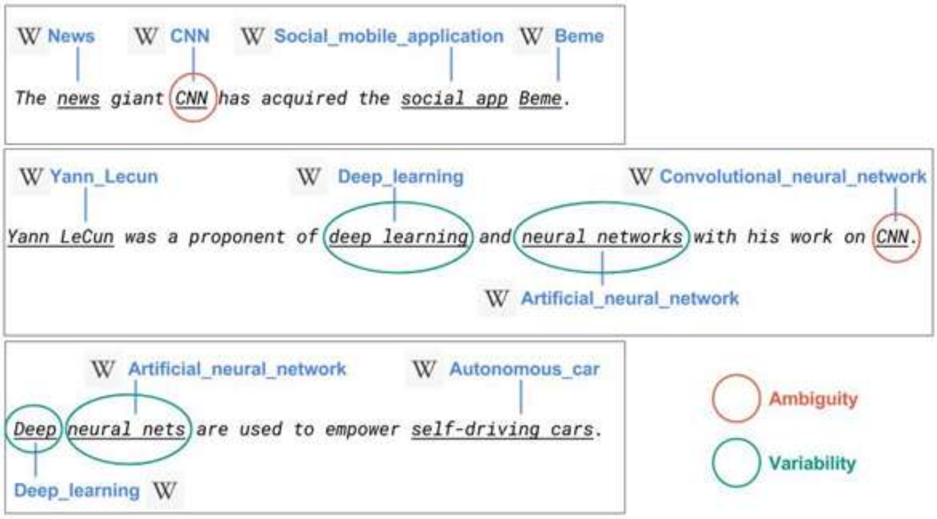

Figure 1.2. Analysis of texts in ambiguity

II. Literature Survey

Named Entity Recognition is an outstanding issue in the field of NLP. Some named entity (NE) taggers like the Stanford Tagger [7] and the Illinois Named Entity Tagger [12] have been appeared to function admirably for appropriately organized sentences. Nonetheless, these NE taggers are probably not going to perform attractively on the inadequate, divided and ungrammatical sentences regularly found in smaller scale posts. Therefore, NE labeling for microposts has risen as a testing research issue. Ritter et al. [14] were among the most punctual to think about NER from tweets. They demonstrate that "the execution of standard NLP devices is seriously corrupted on tweets." Their methodology, in view of Latent Dirichlet Allocation (LDA), uses the Freebase word references, and fundamentally outflanks the Stanford NER framework. Comprehending Micro-posts (#MSM) is a workshop arrangement that began in 2011. It centers on the issue of Information Extraction from small scale posts all in all. A Concept Extraction Challenge (or contest) was organized as a part of #MSM2013. Challenge members were required to accurately distinguish substances having a place with one of four conceivable composes: ‘Organization’, ‘Person’, ‘Miscellaneous’ and ‘Location’. The best test accommodation was by Habib et al. [9]. They utilized a cross breed approach that consolidates Conditional Random Fields (CRF) and Support Vector Machines (SVM) to tag named substances in small scale posts.

2.1. NER on Tweets

Finin et al. (2010) use Amazons Mechanical Turk service 2 and Crowd Flower 3 to annotate named entities in tweets and train a CRF model to evaluate the effectiveness of human labeling. Conversely, our work means to assemble a framework that can consequently distinguish named elements in tweets. To accomplish this, a KNN classifier with a CRF show is joined to use cross tweets data, and the semi-supervised learning is embraced to use unlabeled tweets.

2.2. NER on Non-Tweets

NER has been broadly examined on formal content, for instance, different methodologies have been proposed. For instance, Krupka and Hausman (1998) utilize manual principles to extricate substances of predefined composes; Zhou and Ju (2002) receive Hidden Markov Models while Finkel (2005) utilize CRF to prepare a successive NE labeler, in which the BIO (which means Beginning, the Inside and the Outside of an element, individually) pattern is connected. Different strategies, for example, arrangement in view of Maximum Entropy models and consecutive use of Perceptron or Winnow (Collins, 2002), are additionally polished. The best in class framework, e.g., the Stanford NER, can accomplish a F1 score of more than 92.0% on its test set.

Biomedical NER speaks to a different line of dynamic research. Machine learning based frameworks are normally utilized and outflank the control based frameworks. A best in class biomedical NER framework (Yoshida and Tsujii, 2007) utilizes Syntactic and semantic features, lexical features, orthographic features, such as shallow parsing and part-of-speech (POS).

2.3. Semi-directed Learning for NER

Semi-supervised learning misuses both marked and un-labeled information. It demonstrates helpful when named information is rare and difficult to build while unlabeled information is copious and simple to get to. Bootstrapping is a regular semi-managed learning strategy. It iteratively includes information that has been unhesitatingly named but on the other hand is useful to its preparation set, which is utilized to re-prepare its model. Jiang and Zhai (2007) propose an adjusted bootstrapping calculation and effectively apply it to NER. Their strategy depends on case re-weighting, which permits the little measure of the bootstrapped preparing sets to have an equivalent weight to the expansive source space preparing set. Wu et al. (2009) propose another bootstrapping calculation that chooses connecting occurrences from an unlabeled target area, which are educational about the objective space and are additionally simple to be accurately marked. We embrace bootstrapping also, however utilize human marked tweets as seeds.

III. Comparision Survey Analysis

Table 3.1. Comparision table that Techniques used in various papers

S.No	Paper Name	Author name	Technique Used
1	A framework for benchmarking entity annotation Systems	Marco Cornolti, Paolo Ferragina, Massimiliano Ciaramita	Benchmark Framework
2	A Generative Entity-Mention Model for Linking Entities with Knowledge Base	Xianpei Han and Le Sun	generative probabilistic model

3	Entity Linking on Microblogs with Spatial and Temporal Signals	Yuan Fang and Ming-Wei Chang	A novel entity linking framework
4	Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions	Wei Shen, Jianyong Wang and Jiawei Han	Entity Linking System
5	Exploiting Entity Linking in Queries for Entity Retrieval	Faegheh Hasibi, Krisztian Balog and Svein Erik Bratsberg	Markov Random Field framework and semistructured retrieval
6	Exploiting Wikipedia as External Knowledge for Named Entity Recognition	Jun'ichi Kazama and Kentaro Torisawa	CRF-based NE tagger.
7	Fast and Space-Efficient Entity Linking in Queries	Roi Blanco, Giuseppe Ottaviano, Edgar Meij	Probabilistic model
8	FASTUS: Extracting Information from NaturalLanguage Texts	Jerry R. Hobbs, Douglas Appelt, John Bear, David Israel, Megumi Kameyama, Mark Stickel, and Mabry Tyson	FASTUS System
9	Fine-Grained Entity Recognition	Xiao Ling and Daniel S. Weld	Automatic Labeling Process
10	Fine-Grained Location Extraction from Tweets with Temporal Awareness	Chenliang Li and Aixin Sun	POI inventory and a time-aware POI tagger
11	GERBIL: General Entity Annotator Benchmarking Framework	Ricardo Usbeck, Michael Röder and Axel Cyrille Ngonga Ngomo	GERBIL an evaluation framework
12	Joint Recognition and Linking of Fine-Grained Locations from Tweets	Zongcheng Ji, Aixin Sun, Gao Cong and Jialong Han	Novel Joint Framework and Beam Search Based Algorithm
13	Large-Scale Named Entity Disambiguation Based on Wikipedia Data	Silviu Cucerzan	Named entity disambiguation System
14	Linking Named Entities with Knowledge Base via Semantic Knowledge	Wei Shen, Jianyong Wang, Ping Luo, Min Wang	A Novel Framework LINDEN
15	Link Entities in Web Lists with Knowledge Base	Wei Shen, Jianyong Wang, Ping Luo, Min Wang	LIEGE Framework
16	Named Entity Recognition in Tweets: An Experimental Study	Alan Ritter, Sam Clark, Mausam and Oren Etzioni	A Novel T-NER System
17	Spatial Role Labeling: Towards Extraction of Spatial Relations from Natural Language	Parisa kordjamshidi, Martijn van otterlo and Marie-francine moens	Context-dependent learning techniques
18	TwiNER: Named Entity Recognition in Targeted Twitter Stream	Chenliang Li, Jianshu Weng, QiHe, YuxiaYao, Anwitaman Datta, Aixin Sun, and Bu-Sung Lee	Named Entity Recognition System and TwiNER
19	Re-ranking for Joint Named-Entity Recognition and Linking	Avirup Sil Alexander Yates	Re-Ranking Method
20	Probabilistic Bag-Of-Hyperlinks Model for Entity Linking	Octavian-Eugen Ganea, Marina Ganea, Aurelien Lucchi, Carsten Eickhoff and Thomas Hofmann	light-weight graphical model and Entity disambiguation

Table 3.2. Comparision table of various paper merits and demerits

S.No	Paper Name	Merits	Demerits
1	A framework for benchmarking entity annotation Systems	Comparison of entity annotation system	The human labeled instances in the available datasets over a wide spectrum of entity
2	A Generative Entity-Mention Model for Linking Entities with Knowledge Base	can easily incorporate multiple types of heterogenous entity knowledge	The dependence between entities in the same document are not considered
3	Entity Linking on Microblogs with Spatial and Temporal Signals	Improve entity linking process	Integration of various meta-data are available
4	Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions	Have exploited the geographic aspects of tweets to infer the matches between tweets and restaurants	information extraction and Semantic Web area still need improvements
5	Exploiting Entity Linking in Queries for Entity Retrieval	First attempt of incorporating entity linking into entity retrieval	Retrieval process is not applied for all entity type.
6	Exploiting Wikipedia as External Knowledge for Named Entity Recognition	Improved the accuracy Named Entity Recognition (NER)	In some cases this method incorrectly extracted categories
7	Fast and Space-Efficient Entity Linking in Queries	Our method leverages information from query logs and anchor texts to automatically	The semantic similarity of words are not considered
8	FASTUS: Extracting Information from NaturalLanguage Texts	The system provides a direct link between the texts being analyzed and the data Being extracted	The hard problems cannot be ignored forever, and scientific progress Requires that they be addressed
9	Fine-Grained Entity Recognition	This approach solves a multi-label multi-class Classification problem by adapting a perceptron model	Nose is produced during the distance supervision
10	Fine-Grained Location Extraction from Tweets with Temporal Awareness	This system largely tackles the problem of predominant usage of colloquial language in tweets	context derived from historical tweets from a user is still make some problems
11	GERBIL: General Entity Annotator Benchmarking Framework	It push annotation system developers to better quality and wider use of their frameworks and include the provision of persistent URLs for reproducibility and archiving	The algorithm underlying the web service has Are not stable
12	Joint Recognition and Linking of Fine-Grained Locations from Tweets	This system allows global features to alleviate the error propagation problem	Limited amount of labeled data are only used
13	Large-Scale Named Entity Disambiguation Based on Wikipedia Data	Over news stories and Wikipedia articles this system shows high disambiguation accuracy	This system only works as a Web Browser
14	Linking Named Entities	Name ambiguity,	Still the accuracy rate

	with Knowledge Base via Semantic Knowledge	textual inconsistency, and lack of world knowledge in the knowledge base problems are tackled.	is low
15	Link Entities in Web Lists with Knowledge Base	Enhance link listing process	Table annotation problem occurs
16	Named Entity Recognition in Tweets: An Experimental Study	Improve the performance of Standard NLP tools	Tweets often lack enough context to identify the types of the entities
17	Spatial Role Labeling: Towards Extraction of Spatial Relations from Natural Language	Provides feasible learning of spatial role labeling task	This system only provides single spatial indicators
18	TwiNER: Named Entity Recognition in Targeted Twitter Stream	The error-prone and short nature of Twitter is improved	Entity type classification problems are not included in this process
19	Re-ranking for Joint Named-Entity Recognition and Linking	This model can handle classification of entity links	It requires advanced methods to improve the linking accuracy
20	Probabilistic Bag-Of-Hyperlinks Model for Entity Linking	Improve the performance of entity linking benchmarking collections	Our model considers only pairwise potentials

IV. Conclusion

Our research concludes a survey on comparative study to linking location technique to various mining operations such as lexicon base approach and machine learning based approach that together we have an cross-lingual and cross-domain based methods for some more evaluation metrics and its results. The machine learning methods using Support Vector Machine and Naïve Bayes have very highest accuracy to compare on baseline methods using lexicon models, in some cases we have need to give few effort on human label document methods, more cleaner data give more accurate results to recognition of linking locations. Another sentiment analysis approach called bi-gram model it provides more accuracy to compared mode of studying focal linking locations, it improves accuracy of sentiment analysis classification with various domains of different languages.

References

[1]. M. Cornolti, P. Ferragina, and M. Ciaramita, “A framework for benchmarking entity-annotation systems,” in Proc. 22nd Int. Conf. World Wide Web, 2013, pp. 249–260.

[2]. X. Han and L. Sun, “A generative entity-mention model for linking entities with knowledge base,” in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Human Language Technol., 2011, pp. 945–954.

[3]. Y. Fang and M.-W. Chang, “Entity linking on microblogs with spatial and temporal signals,” Trans. Assoc. Comput. Linguistics, vol. 2, pp. 259–272, 2014.

[4]. W. Shen, J. Wang, and J. Han, “Entity linking with a knowledge base: Issues, techniques, and solutions,” IEEE Trans. Knowl. Data Eng., vol. 27, no. 2, pp. 443–460, Feb. 2015.

[5]. F. Hasibi, K. Balog, and S. E. Bratsberg, “Exploiting entity linking in queries for entity retrieval,” in Proc ACM Int. Conf. Theory Inf. Retrieval, 2016, pp. 209–218.

[6]. J. Kazama and K. Torisawa, “Exploiting wikipedia as external knowledge for named entity recognition,” in Proc. Joint Conf. Empirical Methods Natural Language Process. Comput. Natural Language Learn., 2007, pp. 698–707.

[7]. R. Blanco, G. Ottaviano, and E. Meij, “Fast and space-efficient entity linking for queries,” in Proc. 8th ACM Int. Conf. Web Search Data Mining, 2015, pp. 179–188.

[8]. Jerry R. Hobbs, Douglas Appelt, John Bear, David Israel, and Mabry Tyson “FASTUS: A System for Extracting Information from Natural-Language Text”.

[9]. X. Ling and D. S. Weld, “Fine-grained entity recognition,” in Proc. 26th AAAI Conf. Artif. Intell., 2012, pp. 94–100.

[10]. C. Li and A. Sun, “Fine-grained location extraction from tweets with temporal awareness,” in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 43–52.

[11]. R. Usbeck, et al., “GERBIL: General entity annotator benchmarking framework,” in Proc. 24th Int. Conf. World Wide Web, 2015, pp. 1133–1143.

[12]. Z. Ji, A. Sun, G. Cong, and J. Han, “Joint recognition and linking of fine-grained locations from tweets,” in Proc. 25th Int. Conf. World Wide Web, 2016, pp. 1271–1281.

[13]. S. Cucerzan, “Large-scale named entity disambiguation based on wikipedia data,” in Proc. EMNLP-CoNLL, vol. 7, 2007, pp. 708–716.

[14]. W. Shen, J. Wang, P. Luo, and M. Wang, “Linking named entities in tweets with knowledge base via user interest modeling,” in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 68–76.

[15]. W. Shen, J. Wang, P. Luo, and M. Wang, “Liege:: Link entities in web lists with knowledge base,” in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 1424–1432.

[16]. A. Ritter, S. Clark, O. Etzioni, and Mausam, “Named entity recognition in tweets: An experimental study,” in Proc. Conf. Empirical Methods Natural Language Process., 2011, pp. 1524–1534.
[17]. P. Kordjamshidi, M. van Otterlo, and M. Moens, “Spatial role abeling: Towards extraction of spatial relations from natural language,” ACM Trans. Speech Language Process., vol. 8, no. 3, pp. 4:1–4:36, 2011.
[18]. C. Li, et al., “TwiNER: Named entity recognition in targeted twitter stream,” in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2012, pp. 721–730.
[19]. A. Sil and A. Yates, “Re-ranking for joint named-entity recognition and linking,” in Proc. 22nd ACM Int. Conf. Inf. Knowl. Manag., 2013, pp. 2369–2374.
[20]. O.-E. Ganea, M. Ganea, A. Lucchi, C. Eickhoff, and T. Hofmann, “Probabilistic bag-of-hyperlinks model for entity linking,” in Proc. 25th Int. Conf. World Wide Web, 2016, pp. 927–938.