# **Entity Linking with a Knowledge Base:**

# **Issues, Techniques, and Solutions**

# **Abstract**

The large number of potential applications from bridging Web data with knowledge bases have led to an increase inthe entity linking research. Entity linking is the task to link entity mentions in text with their corresponding entities in a knowledgebase. Potential applications include information extraction, information retrieval, and knowledge base population. However, thistask is challenging due to name variations and entity ambiguity. In this survey, we present a thorough overview and analysisof the main approaches to entity linking, and discuss various applications, the evaluation of entity linking systems, and futuredirections.

Entity linking can facilitate many different taskssuch as knowledge base population, question answering,and information integration. As the worldevolves, new facts are generated and digitally expressedon the Web. Therefore, enriching existingknowledge bases using new facts becomes increasinglyimportant. However, inserting newly extractedknowledge derived from the information extractionsystem into an existing knowledge base inevitablyneeds a system to map an entity mention associatedwith the extracted knowledge to the correspondingentity in the knowledge base. For example, relationextraction is the process of discovering useful relationshipsbetween entities mentioned in text and the extracted relation requires the process ofmapping entities associated with the relation to theknowledge base before it could be populated intothe knowledge base.

**EXISTING SYSTEM**

The information extractionsystem into an existing knowledge base inevitablyneeds a system to map an entity mention associatedwith the extracted knowledge to the correspondingentity in the knowledge base. On the other hand, an entitymention could possibly denote different named entities.For instance, the entity mention “Sun” can referto the star at the center of the Solar System, a multinationalcomputer company, a fictional character named“Sun-Hwa Kwon” on the ABC television series “Lost”or many other entities which can be referred to as“Sun”. An entity linking system has to disambiguatethe entity mention in the textual context and identifythe mapping entity for each entity mention.

**PROPOSED SYSTEM:**

proposed a probabilistic model which unifiesthe entity popularity model with the entity objectmodel to link the named entities in Web text withthe DBLP bibliographic network. We strongly believethat this direction deserves much deeper explorationby researchers.Finally, it is expected that more research or evena better understanding of the entity linking problemmay lead to the emergence of more effective and efficiententity linking systems, as well as improvementsin the areas of information extraction and SemanticWeb.

**MODULE DESCRIPTION:**

# Number of Modules:

After careful analysis the system has been identified to have the following modules:

1. **Entity linking**

**2. knowledge base**

**3. Candidate Entity Ranking.**

**1.Entity linking**

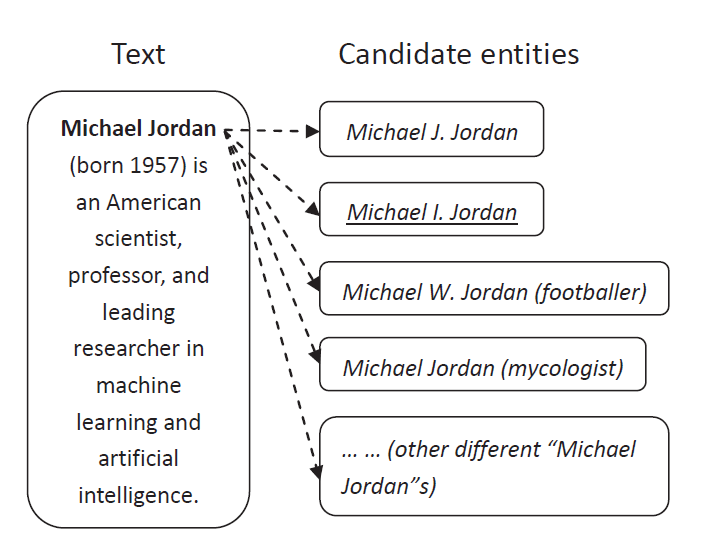
Entity linking can facilitate many different taskssuch as knowledge base population, question answering,and information integration. As the worldevolves, new facts are generated and digitally expressedon the Web. Therefore, enriching existingknowledge bases using new facts becomes increasinglyimportant. However, inserting newly extractedknowledge derived from the information extractionsystem into an existing knowledge base inevitablyneeds a system to map an entity mention associatedwith the extracted knowledge to the correspondingentity in the knowledge base. For example, relationextraction is the process of discovering useful relationshipsbetween entities mentioned in text and the extracted relation requires the process ofmapping entities associated with the relation to theknowledge base before it could be populated intothe knowledge base. Furthermore, a large numberof question answering systems rely on their supportedknowledge bases to give the answer to theuser’s question. To answer the question “What isthe birthdate of the famous basketball player MichaelJordan?”, the system should first leverage the entitylinking technique to map the queried “Michael Jordan”to the NBA player, instead of for example, theBerkeley professor; and then it retrieves the birthdateof the NBA player named “Michael Jordan” from theknowledge base directly. Additionally, entity linkinghelps powerful join and union operations that canintegrate information about entities across differentpages, documents, and sites.The entity linking task is challenging due to namevariations and entity ambiguity.

**2. Knowledge base:**

Given a knowledge base containing a set of entities Eand a text collection in which a set of named entitymentions M are identified in advance, the goal ofentity linking is to map each textual entity mentionm ∈ M to its corresponding entity e ∈ E in theknowledge base. Here, a named entity mention mis a token sequence in text which potentially refersto some named entity and is identified in advance.It is possible that some entity mention in text doesnot have its corresponding entity record in the givenknowledge base. We define this kind of mentions asunlinkable mentions and give NIL as a special labeldenoting “unlinkable”. Therefore, if the matchingentity e for entity mention m does not exist in theknowledge base an entity linking systemshould label m as NIL. For unlinkable mentions, thereare some studies that identify their fine-grained typesfrom the knowledge base which is outof scope for entity linking systems. Entity linking isalso called Named Entity Disambiguation (NED) inthe NLP community. In this paper, we just focus onentity linking for English language, rather than crosslingualentity linking Typically, the task of entity linking is precededby a named entity recognition stage, during whichboundaries of named entities in text are identified.While named entity recognition is not the focus ofthis survey, for the technical details of approachesused in the named entity recognition task, you couldrefer to the survey paper and some specificmethods In addition, there are many publiclyavailable named entity recognition tools, suchas Stanford NER1, OpenNLP2, and LingPipe3. Finkelet al. introduced the approach used in StanfordNER. They leveraged Gibbs sampling augmentan existing Conditional Random Field based systemwith long-distance dependency models, enforcing labelconsistency and extraction template consistency.

**3.Candidate Entity Ranking**

In most cases, the size of the candidate entityset Em is larger than one. Researchers leveragedifferent kinds of evidence to rank the candidateentities in Em and try to find the entity e ∈ Emwhich is the most likely link for mention m. InSection we will review the main techniquesused in this ranking process, including supervisedranking methods and To deal with the problem of predicting unlinkablementions, some work leverages this module tovalidate whether the top-ranked entity identifiedin the Candidate Entity Ranking module is thetarget entity for mention m. Otherwise, they returnNIL for mention m. In, we willgive an overview of the main approaches forpredicting unlinkable mentions.



**System Configuration:**

**HARDWARE REQUIREMENTS**:

Hardware - Pentium

Speed - 1.1 GHz

RAM - 1GB

Hard Disk - 20 GB

Floppy Drive - 1.44 MB

Key Board - Standard Windows Keyboard

Mouse - Two or Three Button Mouse

Monitor - SVGA

**SOFTWARE REQUIREMENTS**:

Operating System : Windows

Technology : Java and J2EE

Web Technologies : Html, JavaScript, CSS

IDE : My Eclipse

Web Server : Tomcat

Tool kit : Android Phone

Database : My SQL

Java Version : J2SDK1.5