WiiCluster: a Platform for Wikipedia Infobox Generation
Kezun Zhang, Yanghua Xiao, Hanghang Tong, Haixun Wang and Wei Wang

**Introduction**

- **Generating Infobox from Wikipedia**
  - Structured information in Wikipedia such as table, image, text, citation, etc., all of which are the targets of information extraction and are machine processible. More importantly, many entities are associated with an Infobox which consists of a set of (property, value) pairs about the entities. But current Infobox in Wikipedia is incomplete. This paper proposes a novel method to generate structural information from Wikipedia to supplement existing Infobox.

- **Linked Entity Based Method**
  - The key idea of this paper is the following: If we can summarize the relationship between the entity and its linked entities, then we immediately harvest some of the most important information about the entity.

**How to accurately summarize linked entities?**
- **Clustering and tagging**
  - Clustering. We make use of the Wikipedia category system, which select categories from the category system as the feature of the entity and then weight the feature. But the existing category system is not perfect, because some hypernyms do not describe the IS-A relationship, which may enlarge error in clustering. So we use IS-A relation preserved method to select and weight feature.
  - **Maximal ξ-LCA(Least Common Ancestor) tagging method**

**How to efficiently extract knowledge for all articles?**
- We use cluster reuse method to accelerate scalable clustering tasks. Since many articles share same linked entities.

**Flowchart**

- **WiiCluster**
  - A Web Interface is developed for searching and browsing the clusters of the linked entities of an entity. The clusters are grouped by the cluster tag.

**Implements**

- **How to accurately summarize linked entities?**
  - Clustering and tagging
    - Clustering. We make use of the Wikipedia category system, which select categories from the category system as the feature of the entity and then weight the feature. But the existing category system is not perfect, because some hypernyms do not describe the IS-A relationship, which may enlarge error in clustering. So we use IS-A relation preserved method to select and weight feature.

**Demonstration**

- We present two different models to visualize the clusters: aggregated graph model in Fig. 4 and the tree model in Fig. 5.