H22年 1 月 15 日

Electronics and Information Engineering Department	ID	055401
Name	MD. HANIF SEDDIQUI	

Advisor	MASAKI AONO

Title	**Ontology Alignment and Its Application to Knowledge Engineering**

(800 words)

In this thesis, we focus on ontology alignment of semantic web and its application to instance matching, multimedia resource integration, and to text classification. It has been a formidable task to achieve efficiency and scalability in aligning two massive and conceptually similar ontologies. A straightforward approach to the process entails an $O(N^2)$ computation, where N denotes the average number of concepts in each ontology. Our Anchor-Flood algorithm of ontology alignment, boasting of $O(N\ log(N))$ computation on the average, starts off with an *anchor* (a pair of "look-alike" concepts from each ontology), gradually collecting blocks of neighboring concepts across ontologies. We compare and compute similarity among concepts of the collected blocks across ontologies to produce more aligned pairs. Considering each of the newly found aligned pairs as an anchor, our algorithm repeats the similarity computation iteratively until the algorithm meets that *"either all the collected concepts are explored, or no new aligned pair is found"*. Eventually it outputs an alignment between concepts and properties within semantically connected subsets of two entire graphs, which we call *segmented alignment*. We have achieved the best runtime and scalability in Ontology Alignment Evaluation Initiative (OAEI), 2008 and 2009 campaign. Our further experimental results reveal the fact that the larger block often encounters some non-relevant concepts. Therefore, we propose a metric of Intrinsic Information Content (IIC) to measure semantic similarity among concepts of a block of neighboring concepts of an ontology. Our proposed IIC measures semantic similarity by the concept-concept, concept-property and concept-instance-property relations of an ontology. Our experiments show the effectiveness of our proposed metric of measuring IIC.

Furthermore, we apply our ontology alignment algorithm to knowledge engineering of instance matching, multimedia resource integration and text classification. We propose an efficient algorithm for ontology instance matching by considering the semantic links associated to a particular instance. An instance is specified explicitly with the help of linked concepts, properties and their values in a neighborhood. We refer these neighboring links of an instance as a *"Semantic Link Cloud" (SLC)* in a collection. We compare SLCs to align ontology instances across heterogeneous sources of ontology knowledge base. Our algorithm shows its strength over various transformations: value transformation, logical and structural transformation. In addition, we use our ontology alignment and the instance matching algorithm to integrate MPEG-7 based heterogeneous multimedia resources efficiently. Moreover, we introduce a method of ontology driven text classification that retrieves correspondence between research abstract and patent categories, i.e. International Patent Classification (IPC), organized in a taxonomy. In our proposed method, we associate each IPC to a set of related features and their weighting factors extracting from the pre-classified huge patent documents by well-known IR techniques. We extract features from a research abstract and incorporate a classical classifier to produce probable IPCs. Then we look up the features at each of the probable IPC along with its neighbors to measure the degree of relevance. Our experimental results show the efficiency of our method.

The detailed experiments prove the efficiency, and scalability of our Anchor-Flood algorithm and its applicability to various research domains of knowledge engineering.