

Building an Event Ontology for Historical Domain to Support Semantic Document Retrieval

Fatihah Ramli ^{#,*} and Shahrul Azman Mohd Noah ^{*}

[#] Department of Information Systems, Faculty of Computer Science & Information Technology, Universiti Malaysia Sarawak (UNIMAS),
94300 Kota Samarahan, Sarawak, Malaysia
E-mail: rfatihah@unimas.my

^{*} Knowledge Technology Research Group, Faculty Information Science and Technology, Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor, Malaysia
E-mail: shahrul@ukm.edu.my

Abstract— In the past years, there has been increasing concern on ontology for its ability to explain data semantics in the usual manner independent of the data source characteristics, providing a schema that allows interchanging data between heterogeneous information systems and users. The ontology development in some areas is not expected due to a large amount of information, particularly in history, leading its semantic impossible. Several works have been designed to improve the technological aspects of ontology, such as the representation of language and inference mechanisms, and less attention has been paid to practical results development of application methods. This paper presents a discussion on the experience and processes during ontology building in history: historical documents retrieval based on the event.

Keywords— ontology; history; event; historical document; historical domain; semantic retrieval

I. INTRODUCTION

Ontology has received recognition from the academy and industry in various fields [1]. The definitions of ontology vary according to the fields and applications. In information science, ontology can be defined as a dictionary of terms formulated in a canonical syntax and with commonly accepted definitions designed to yield a lexical or taxonomical framework for knowledge representation, which can be shared by different information systems communities[2], [3]. Thus, ontology is said to be a representation of the things that exist within a particular domain of reality such as medicine, geography, finance, or history. The development of the ontology for these specific domains is meant to support the implementation of intelligent applications such as decision support systems[4], recommender systems[5] and semantic search[5],[6].

One of the domains receiving great attention recently is history[7], [8], which may be due to increasingly available digitised historical documents and artefacts to the public. History can be referred to as a period of time after writing was invented. It is a field of research that uses narrative to examine and analyse the sequence of events, and it sometimes attempts to investigate the patterns of cause and

effect that determine events objectively. Research on managing historical documents involves finding, using and correlating the documents in order to communicate an understanding of past events. Historical documents can be defined as those that keep the information related with time instant at which the documents were published at the same time that is still useful in the future [9]. According to Elena [10], [11], within the context of the historical archive, historians employ their knowledge, experience and intuition to decide on the information that they need to find and study; and attempt to locate sources that contain the information. The results from Elena [10] obviously stated that historians need historical sources repositories and building tools to enable them to access comprehensive information rapidly. Among the most important information for them is the event. Questions such as: *When did the specific event occur?*, *What are the relations among events?*, *Who were involved?* and *List the chronological of specific events*. An obvious way to retrieve such information from large repositories is via information retrieval (IR) systems, or commercially known as search engines. IR is a field concerned with the structure, analysis, organisation, storage, search, and retrieval of information" [12]. According to [13], the need of IR research areas led to the creation of semantic web. However,