A Collective Intelligence Framework for Supporting Universal Information Agents

Marut Buranarach, Alisa Kongthon, Chatchawal Sangkeettrakarn, Sarawoot Kongyoung, and Supon Klaithin

Human Language Technology Laboratory
National Electronics and Computer Technology Center (NECTEC)
112 Thailand Science Park, Phahon Yothin Rd.
Klong Luang, Pathumthani, Thailand 12120
{marut.buranarach, alisa.kongthon, chatchawal.sangkeettrakarn, sarawoot.kongyoung, supon.klaithin}@nectec.or.th

Abstract. Domain knowledge is important for an agent system both in terms of providing background knowledge in agent conversation and for a knowledge question-answering service. In this paper, we present our framework for constructing a knowledge base for the ABDUL system, a universal information agent, aiming to provide a universal access to information and knowledge. Our knowledge construction approach is based on collective intelligence (CI) applied to the user community. We discuss existing approaches and our design rational for CI-based knowledge construction tool based on the Semantic Web framework.

Keywords: Agent-Based System, Knowledge Construction, Semantic Web

1 Introduction

Information overload has become an inevitable problem in our modern life. To overcome such problem, users typically rely on search engines. Though these engines are excellent tools, they rather yield the list of documents that matches with users’ search criteria than the information that they really need. As a result we have developed an interactive conversational agent called ABDUL\(^1\) (Artificial BudDy U Love) as an online information service which currently provides access to Thai linguistic and information resources [1]. ABDUL allows users to connect to the service from various devices and platforms via Instant-Messaging-based protocol. User interaction is similar to that of a question answering (QA) system that allows users to retrieve answers to questions posed in natural language form. Based on initial statistics, our service has shown to be promising in terms of new user registration and service usage [2].

ABDUL is now served as a language and information assistant. In term of language assistant, ABDUL can help translate vocabularies (Thai to English and

\(^{1}\) http://www.abdul.in.th/
English to Thai) and sentences (English to Thai). As an information assistant, users can ask for general information such as weather condition, stock price, gas price, traffic condition, among others. ABDUL will then extract information from the corresponding public Web sites to answer such queries. Hence, our current system relies mainly on existing generic information available online. However, to be able to respond to user’s questions which require more specific domain knowledge, we need to identify such knowledge reference sources (if any) or else we have to create such knowledge for the system.

Domain knowledge is important for an agent system [3] both in terms of providing background knowledge in agent conversation and for a knowledge QA service. Traditional knowledge base construction usually relies on few experts in creating a knowledge base. Recent advances in communication and Internet technology have enabled collaborative environment for constructing knowledge bases. The Semantic Web initiative can be considered the largest-scale effort in creating knowledge base for intelligent agents. Success cases of the Web 2.0, or Social Web paradigm, such as Wikipedia, YouTube, and Flickr, have demonstrated potentials of collective intelligence (CI). Thus, recent trends in the Semantic Web research involve how to utilize CI in building large-scale knowledge bases. These efforts include the Semantic MediaWiki [4], DBPedia [5] and Freebase [6] projects. They attempt to transform and store user-generated content into machine-understandable formats, i.e. RDF, OWL, that constitute knowledge for agent systems.

In this paper, we focus on design of a framework for supporting knowledge construction for ABDUL system using CI-based approaches. Architecture and mechanisms that allow the user community to contribute individual knowledge to the system are presented. User contributions can be either conversation patterns [7] or domain knowledge. Finally, we present some rational and criteria for the design of our CI-based knowledge construction tool.

2 CI-Based Knowledge Construction

The Web 2.0 paradigm, also known as the Social Web, has created a large growing amount of user-generated content. Large-scale Web sites such as Wikipedia, LinkedIn, YouTube, Del.icio.us and FaceBook, rely mainly on a huge amount of content contributed by their users. When a critical mass of participation is achieved, user-generated content becomes collective intelligence. Emerging Web 3.0 paradigm, based on the Semantic Web technologies such as RDF, OWL and SPARQL, offers powerful mechanisms for data organization and retrieval. It advocates creation and management of machine-understandable data formats. While the Social Web offers a powerful mechanism for content generation, the Semantic Web complements it in terms of content organization and intelligent retrieval.

One major goal of the Semantic Web is to create large-scale knowledge base for intelligent agents. One of the challenges is how knowledge creation and acquisition can be achieved in a large scale. Success cases of the Web 2.0 paradigm have demonstrated potentials of collective intelligence. Recent trends in the Semantic Web research attempt to utilize collective intelligence in knowledge construction. A
summary of CI-based knowledge construction tools is shown in Fig. 1. The classification is based on whether an approach is based on Wiki-based tools and data and degree of underlying structure in the created data.

<table>
<thead>
<tr>
<th></th>
<th>Unstructured/ Semi-structured</th>
<th>Structured</th>
<th>Structured with Semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wiki-based</strong></td>
<td>Ex. Wikipedia articles</td>
<td>Ex. Wikipedia infobox</td>
<td>Ex. Semantic Wiki</td>
</tr>
<tr>
<td><strong>non Wiki-based</strong></td>
<td>Ex. Webboard, Blogs, User-generated content (UGC)</td>
<td>Ex. Freebase</td>
<td>Ex. Collaborative ontology development environment</td>
</tr>
</tbody>
</table>

* requires data transformation

Fig. 1. Classification of CI-based knowledge construction tools

In assessing a CI-based knowledge construction tool, some main criteria that should be taken into consideration include:

- User's ease in creating the data. Users who are domain experts are not always computer experts. Thus, CI-based tools should be easy to use to promote content creation.
- Degree of effort required in transforming user data into machine understandable formats. Conversion of user-generated data into machine-
understandable formats should be possible and should not require too much effort.

- Level of control imposed on the users. User should have some freedom and flexibility in creating data and modifying structure of the data.
- Level of semantics underlying in the created data. The created data should contain some degree of structure and semantics that provide basis for knowledge base construction.

Fig. 2 shows a comparison of different CI-based knowledge construction tools. The tools are compared in terms of user's ease in creating the data, the amount of effort required in transforming user data into machine-understandable formats, level of semantics underlying in the created data, and level of control imposed on the users.

3. The Proposed CI Framework for ABDUL System

3.1 Design Architecture

To create knowledge for ABDUL system, we propose a framework based on collective intelligence approach as illustrated in Fig. 3. Our framework consists of three main components which can be explained in details as follows:

![Fig. 3. The proposed CI framework for ABDUL system](image-url)
• **Individual Knowledge**: In this stage, each user can register to our system and construct his/her profile. Our system can enable users to create both content for topic of their interests (i.e., domain knowledge) and conversational patterns. We apply Artificial Intelligence Markup Language (AIMA) [8], a markup language for conversation agent construction defined based on XML, in creating conversation templates for ABDUL. For the domain knowledge, we plan to store the information in RDF form, which is the standard data format for the Semantic Web. Once users finish creating their knowledge content, they can choose to share it in the user community.

• **Global Knowledge**: If users decide to share their knowledge in the community, such topics and content will then need to be verified. Our verification process is based on administrator approval and user voting. In order to be distributed to the community, topics and content have to be based on fact not personal opinion.

• **Agent Conversational Modules**: Once global conversation patterns and domain knowledge are constructed, they can be used by ABDUL conversational modules such as WebChat, MobileChat or MSNChat to interact with the users.

### 3.2 Knowledge Base Construction Tool

In the case of ABDUL user community, users are varied in terms of technological knowledge and skills. Tools such as Semantic Wiki and ontology editors require some advanced training and may not be most suitable for the user community. Typical Wiki-based tools are most friendly in allowing the users to contribute their knowledge. However without some control imposed on the users's created data, the task of transforming the data to agent knowledge would be too demanding.

Our CI initiative adopts a similar approach to Wikipedia infobox and Freebase which attempt to balance between user’s freedom and semantics capturing. Our goal is to maximize user's freedom and ease of creation while maintaining high degree of underlying structure and semantics in the created data. In order to achieve this, we apply Wiki tool and template [9] in creating structured data. Fig. 4 shows an initial experiment with the approach.

![Fig. 4](image-url)  
*Fig. 4. Applying Wiki tool and template for agent knowledge construction*
In this example, user created a wiki page containing facts about a play in terms of attribute names and values, such as play title, actors, release year etc. Hence, the created data can be possibly transformed into RDF triples. User's degree of restrictions and RDF transformation process will be further explored and evaluated in our future work.

4 Conclusion

In this paper, we describe a universal information agent currently providing a unified access to diverse information resources. We present our framework for supporting the user community to contribute their knowledge to the system. The framework supports agent knowledge base construction based on collective intelligence. We discuss existing approaches and our design rational for a CI-based knowledge construction tool based on the Semantic Web framework. Further exploration and evaluation will be required. We believe that the Semantic Web framework applied to collective intelligence will provide an effective basis for creating large-scale knowledge base for the universal information agent framework.

References