Towards Automatic Discovery of Web Portals
Semantic Description of Web Portal Capabilities

Haibo Yu¹, Tsunenori Mine², and Makoto Amamiya²

¹ Graduate School of Information Science and Electrical Engineering,
Kyushu University,
6-1 Kasuga-koen, Kasuga, Fukuoka 816-8580, Japan
yu@al.is.kyushu-u.ac.jp

² Faculty of Information Science and Electrical Engineering,
Kyushu University,
6-1 Kasuga-koen, Kasuga, Fukuoka 816-8580, Japan
{mine, amamiya}@al.is.kyushu-u.ac.jp

Abstract. Due to the problem of information overload, locating relevant
Web portals precisely based on user requirements is quite an essential
task. As the need for application-to-application communication and in-
teroperability grows, providing Web portal services that satisfy human
as well as machine requirements is becoming a new challenge for Web
portals. However, a Web portal capability expressing mechanism, which
enables the precise location of Web portals as well as the automated dis-
covery and invocation of Web portal services, is lacking. In this paper,
we investigate how to incorporate Semantic Web technology with Web
service technologies to describe the capabilities of Web portals. We also
discuss the possibilities of using these descriptions for discovering and
using the distributed existing portal resources.

1 Introduction

Web portals are information rich sites that gather a variety of useful informa-
tion from different resources into a single “one stop” Web page and provide it
in a compact and easily consumable form to an end-user [1]. However, locating
relevant Web portals is quite a challenging task because of the problem of in-
formation overload. It caused us to reconsider user requirements on the Web.
The following questions come to mind: (1) Are there any better ways to locate
relevant Web portal resources precisely based on users’ requirements? (2) Is it
possible to make use of the relevant heterogeneous Web portal resources auto-
matically no matter what framework they are based on? (3) Furthermore, is it
possible to build a user’s own personalized information warehouse (MyPortal)
on her/his computer with these Web portal resources and use it conveniently,
even sharing it with other people? In this paper, we are trying to answer the
first question, as a preliminary step towards answering the latter two.

Locating Web portal resources should be based on a match between user
requirements and Web portal capabilities. This requires a mechanism for ex-
pressing the capabilities of Web portals. Currently, there are some standards used for describing the capabilities of Web sites [1] [2], but they are generally used for the aggregation of Web sites or portlets into a Web portal, not for application computing purposes. So an explicit description of Web portal capabilities, which can support automatic discovery of the Web portal as well as its services, is lacking.

Semantic Web [3] is an evolving technology which aims to tackle the information overload problem of the current Web. In the Semantic Web, the information is given well-defined meaning, better enabling computers and people to work in cooperation.

Web service mechanisms provide a good solution for application interoperability between heterogeneous environments. They are standard programmatic interfaces between applications that provide a new model which enables Web sites to exchange dynamic information on demand.

In this paper, we investigate how to incorporate Semantic Web technology with Web service technologies to support the description of Web portal capabilities, trying to enable the precise location of relevant distributed existing portal resources and their maximum reuse. The advantage of our approach is that we provide a mechanism for describing Web portal capabilities that not only enables precise and automatic discovery, but also enables the application to use the Web portal resources after they are located. Since we use standard ontology language and Web service technology, common existing applications, tools and resources can be used.

The rest of the paper is organized as follows: Section 2 briefly describes the basic technologies this research is concerned with. Section 3 describes a mechanism for the description of Web portal capabilities. Section 4 examines the semantic matching algorithm. In section 5 we discuss the possibilities of using these descriptions for discovering portal resources, to enable the maximum reuse of existing distributed portal resources. Related work is discussed in section 6 and the concluding remarks will be summarized in section 7.

2 The Basic Technologies

We next briefly introduce some of the technologies this research is concerned with.

2.1 Semantic Web

The Semantic Web [3] is an extension of the current Web. It is trying to change the current Web into a huge knowledge base with well-defined meaningful data enabling machines to cooperate with people to tackle the problem of information overload.

RDF (Resource Description Framework) [4] is a metadata modeling language recommended by W3C (World Wide Web Consortium). It provides a common framework for expressing information so it can be exchanged between applications without loss of meaning. It uses XML as an interchange syntax.