Knowledge Repositories

Collecting Japanese Terms used in Agricultural Programs

Takuji Kiura

National Agricultural Research Center,
National Agriculture & Food Research Organization
Tsukuba, Ibaraki, Japan
kiura@affrc.go.jp

Tonokazu Yoshida

WeNARC,
National Agriculture & Food Research Organization
Fukuyama, Hiroshima, Japan
jones@affrc.go.jp

Masaki Omine

KONARC,
National Agriculture & Food Research Organization
Miyakonojyo, Miyazaki, Japan
omine@affrc.go.jp

Abstract—**To develop agricultural filed data exchange format, we started collecting Japanese terms and meanings, used in Japanese agricultural applications and XML schema, employing MediaWiki with Semantic MediaWiki Extensions. We defined properties to represent the data schema of each application, and properties to show relations among terms we collect. We started to collect terms and meaning used in schema of PMS and BIX-pp, and found that the usage of an application or XML data format is strongly depends on how he/she use it. And it is not sufficient to collect terms and meaning used in schema to exchange data each other. We should collect terms and its meaning from actual data files of uses, and data exchange depends on the case.**

Field Data; Data Modeling; Data Exchange Format

I. INTRODUCTION

These days, many agricultural application programs are developed and used in Japan. Those handled same kinds of data in their ways. It meaning that treating data without standard data format, standard data exchange format, and/or standard application program interface (API) for exchanging data. Hence users can not exchange data with others, and move one application to another easily. For researchers it is a hard work to integrate data from those applications.

To approach this issue, we start to collect terms and meanings used in applications and data format for Agriculture. This year we focus on terms and meanings appeared in data schema used in applications. If a term used in an application schema has the same meaning of a term used in another application, we can easily integrate data stored in those columns or xml tags.

II. METHOD

We work in the different Research Centers in Japan and need a collaborative work environment such as Wikipedia. MediaWiki is freely available and use many projects in the world (http://www.mediawiki.org/wiki/MediaWiki.) We also need the function to express the relation between terms we collect. The Semantic MediaWiki Extension (SMW) is a free extension package for MediaWiki, and we can add semantic annotation to wiki pages in simple manner (http://semantic-mediawiki.org/wiki/Semantic_MediaWiki.) Using SMW, we can describe the relation between wiki pages and properties of pages. Nagai et al. also using MediaWiki with SMW in their "Network Dictionaries" [1]. One of authors doing collaborative works with him under the project "Data Integration and Analysis System (DIAS, http://www.editoria.u-tokyo.ac.jp/dias/english/index.html)" and we want to share terms with his group. Hence we decided to use MediaWiki with SMW.

We generated a virtual machine using VMware Workstation Version 7.1 and installed Ubuntu 10.04 Server Edition [2]. If more resources are required, we can add resources to the virtual machine, or we can move the virtual machine to another virtual host. Ubuntu 10.04 is the latest Long Term Support version and free. It has MediaWiki package and SMW package in its universe repository. We install those packages and set up MediaWiki with SMW. Consequently, we use MediaWiki version 1.15.1 and SMW version 1.4.0.

We added properties for applications used some relational database management systems (RDBMS) and XML schema. (Table I.) To avoid collisions of page name, we use java class style naming methods. For example, a column "temp" in a table "datatable" in the application "hoge", developed or sold by example.com, we create a page "com.example.hoge.databable.temp" and describe the meaning in the application. We also defined templates to make page description easier.

For our experiments, terms appeared in two systems (PMS and BIX-pp), are described. PMS is a planning and management support system for agriculture developed by Dr. Tomokazu Yoshida, one of authors, and uses RDBMS (http://www.aginfo.jp/PMS). BIX-pp (BioInformation Exchange foro Plant Production) is a data exchange XML format developed by Dr. Takehiko Hoshi [3]. We are collaborating with him in another project. Thus we are familiar to both systems. Therefore we chose those for experimental data input to wiki system.

TABLE I.   PROPERTIES DEFINED

Property name	Schema Type	Meaning
has table	RDBMS	Page, corresponding database or system, has a table name of another page
is a table of	RDBMS	Page is a table name of page corresponding to database or system. Reverse relation of "has table".
has column	RDBMS	Page, corresponding table, has a column name of another page.
is a column of	RDBMS	Page, corresponding to column, appears in the table
has value	RDBMS	Page corresponding to column has value of other page
is a value of	RDBMS	Page corresponding to value appears in another page corresponding column of a table. Reverse relation of "has value"
refers to	RDBMS	Foreign key
has element	XML	Page, corresponding to XSD file, has a element page
is a element of	XML	Reverse relation of "has element".
has simpleType	XML	Page, corresponding to XSD file, has a simpleType page
is a simpleType of	XML	Reverse relation of "has simpleType"
has complexType	XML	Page, corresponding to XSD file, has a complexType page
is a complexType of	XML	Reverse relation of "has complexType"
xsd:ref	XML	Refers another element page
xsd:base	XML	Element is extended from another

(schema-related properties only)

III.   RESULT AND DISSCUSSION

Note that experimental data input is still undergoing. Results may be changed.
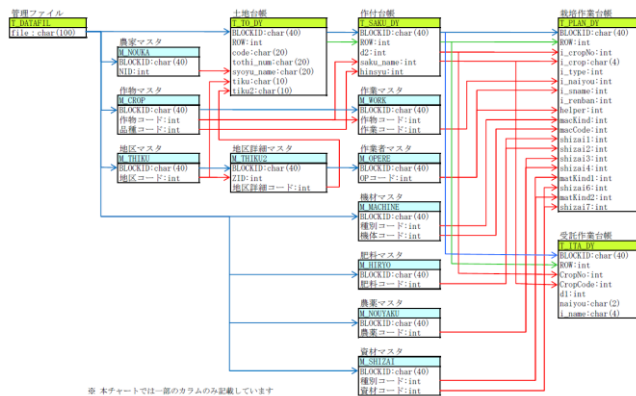
A.   PMS

FIGURE 1.   MAIN PART SCHEMA OF PMS

We finished describing main part of PMS schema (Fig. 1). In this experiment, we can create pages those represent the data schema of RDBMS.   Fig. 2 shows pages,

jp.aginfo.pms (PMS), jp.aginfo.pms.m_crop (crop master table,) and jp.aginfo.pms.m_crop.品種名 (column to store variety of crop,) in our wiki system. But the property "refers to" is not enough to describe complex foreign key constraints, not used in PMS.

FIGURE 2.   WIKI PAGES FOR PMS

Some table should be prepared by users before use. For example, users must be define which kind of variety they want to use in jp.aginfo.pms.m_crop.品種名. This feature of PMS is convenient for users, because it brings users

flexibilities to use PMS for their own purpose. Some uses jp.aginfo.pms.m_crop.品種名 for variety and cropping method. Jp.aginfo.pms.m crop.品種名.有機コシヒカリ means that variety is "KoshiHikari", one of the most popular variety of rice in Japan, and be cropped with organic fertilizers and no agricultural chemical. Another uses jp.aginfo.pms.m_crop.品種名 for methods or places such as Jp.aginfo.pms.m crop.品種名.水張り保全, means irrigated and preserved for next cultivation.

Accordingly, collecting terms and meanings in PMS schema is not enough to exchange data with other systems and with other users of PMS. We should investigate into use cases, and the way to exchange data is depending on each case.

B. BIX-pp

We described the BIX-pp XML schema file [4] using our wiki system referring to BIX-pp example xml file [5]. In this experiment, we found followings.

- "Altitude", "Area", "Latitude", and "Longitude" are defined without "Unit". According the example, "Altitude" unit is "meter", Area" unit is "square meter", "Latitude" and "Longitude" units are "degree". Fig. 3 shows the definition of "Altitude" in BIX-pp schema. If a user doesn't refer to the example, he/she may use other units for those elements. Hence we should check the data carefully to exchange data.

```
<xsd:element name="Altitude" type="xsd:double"/>
```

FIGURE 3.   DEFINITION OF "ALTITUDE" IN BIX-PP

- "DataItem" doesn't have predefined IDs. Fig. 4 shows a "DataItem" part of the example file [5]. This declares "DataItem" having "Index" 1 is used for "Daily Average of Air Temperature" with unit "C", may be "degree Celsius". So users can record any data type using "DataItem" tag in BIX-pp XML file. From data exchange point of view, "Name" of "Daily Average of Air Temperature" is used without any reference. Other may wonder that "Daily Average of Air Temperature" is the same meaning with "Daily Average of Air Temperature" in WMO vocabulary. Someone may use "Daily Average of Air Temperature" another meaning. Therefore, data integration with BIX-pp data among the BIX-pp users is hard work. To integrate the values stored as "DataItem" with other data, we should investigate each BIX-pp XML files in detail.

```
<DataItem>
<!-- Index:1 には、日平均気温を割り当てました -->
<Index>1</Index>
<Name>Daily Average of Air Temperature</Name>
<Unit>C</Unit>
</DataItem>
```

FIGURE 4.   USAGE OF "DATAITEM" IN THE EXAMPLE FILE

C. Standard of Agricultural Vocabulary

If a Japanese Agricultural Vocabulary Standard exists on the World Wide Web, many agricultural applications can be refer the terms in it. Data exchange and integration became easy using the vocabulary references. Food and Agriculture Organization (FAO) have been developing the Concept Server [6] and AGROVOC [7]. Those linguistic recourses can be used for standard agricultural vocabulary. The Ministry of Agriculture, Forestry, and Fisheries, Japan has made a Japanese translation of AGROVOC. Takezaki et al. report that AGROVOC is not enough to retrieve agricultural academic papers efficiently and has developed the Japanese Agricultural Thesaurus (JAT) adding Japanese specific terms to the Japanese translation of AGROVOC [8]. Unfortunately, the explanations of terms are missing in JAT. Addition of term description is desired for JAT to use Japanese Standard of Agricultural Vocabulary. Without standard, data exchange and integration may be complicated tasks.

D. Evaluation of Wiki

Using MediaWiki with SMW, we can describe terms, appeared in two types of agricultural data sources, PMS and BIX-pp, and can represent those schema in wiki system. Our experiment, on PMS and BIX-pp, reveals that some schema of Japanese agricultural application or data exchange format is not well defined, and use case investigations are required to exchange and/or integrate agricultural data. MediaWiki with SMW is flexible for describing use cases to add new properties for use cases. We believe that we can achieve our subject using this system. MediaWiki and SMW can be used other languages, and freely available, therefore others can test our method using another language without charge.

IV.   FUTURE PLAN

We will review the properties we introduced MediaWki with SMW to adapt other types of data models. Collecting terms and its meaning is continued and extended other programs or data formats for agriculture in Japan. We will collect use cases of those programs and formats and describe the meanings using wiki. Then we would like to define our standard Japanese vocabulary for agriculture in this wiki system.

Then we will do data modeling for define our XML format extending BIX-pp to exchange agricultural data. Wolfert et al. already started developing the agricultural data exchange format "agriXchange" [9]. But, we can't find the details of "agriXchange" on "agriXchange" web site

(http://www.agrixchange.eu/). We should inspect it and make our format compatible with it. Finally we will define a service API to exchange data from various agricultural applications to others using standard data format.

V. ACCKNOWLEGEMENT

REFERENCES

[1] M. Nagai, M. Ono, and R. Shibasaki, "Semantic Network Dictionary for Ontological Information with Wiki," posted August 6, 2010, http://www.earthzine.org/2010/08/06/semantic-network-dictionary-for-ontological-information-with-wiki/ (browsed on Sep. 1, 2010)

[2] Canonical Ltd., Ubuntu Server Edition Available: http://www.ubuntu.com/server/get-ubuntu/download

[3] T. Hoshi, "Promoting encapsulation of plant producing information" http://www.agmodel.net/bix-pp/index.htm (browsed on Sep. 1, 2010)

[4] T. Hoshi, "Definition the BIX-pp schema (DTD) by XML schema" Available: http://www.agmodel.net/bix-pp/XMLSchema/BIX-pp.xsd (browsed on Sep. 1, 2010)

[5] T. Hoshi, "An example of the description of BIX-pp file" Available: http://www.agmodel.net/bix-pp/XMLSchema/sample-j.xml (browsed on Sep. 1, 2010)

[6] FAO, AGROVOC Concept Server http://aims.fao.org/website/The-Concept-Server/sub (browsed on Sep. 1, 2010)

[7] FAO, AGROVOC Thesaurus. Available: http://aims.fao.org/website/AGROVOC-Thesaurus/ (browsed on Sep. 1, 2010)

[8] A.Takesaki, M. Saito, and A. Okabe, "Development of Japanese Agricultural Glossary for Efficient Searching", Agricultural Information Research, Vol. 17 (1), 2008, pp. 42-49 [Japanese, english abstract avaibale: Agricultural Information Research, Vol. 17 (1), 2008, p. 49]

[9] J. Wolfert, C. N. Verddow, C. M. Vreloop, and A. J. M. Beulens, "Organizing information integration in agri-food – A mothod based on a service-oriented architecture and libing lav approach", Computers and Electoronics in Aguriculure, vol.70, 2010, pp. 389-405.