DISTRIBUTION OF DATA ON THE WEB

Luminiţa Pătraşcu[1]

Abstract

Data are most often represented [1] in tabular form, where each row represents a few records that are described and each column represents some properties of these items. Cells in a table have particular values for these properties. This article shows a sample of data about a school, completed during the school-leaving examination, 2017. We analyze several different strategies on how these data can be distributed on the web. In all these strategies, some data will be represented on a computer while other parts will be represented on another computer.

Keywords: school leaving-examination, data distribution of Web, strategies for the distribution of data on Web, distributed data, RDF solution, semantic web

The distribution on Web for data reffers to School-leaving examination 2017

Figure 8-1 illustrates a strategy for disseminating information on several computers. Each computer on the network is responsible for maintaining information on the completion of one or more rows in the table. Any query about a particular entity can be honored by the computer that is stored row. A computer is responsible for the information about the school to support students written exam E-school leaving exam and graduation year.

This distribution solution offers considerable flexibility, since the computer can share the load of representative information about a few specific situations [1]. But because this is a distributed representation of data, a few operations coordination between servers are neccesary. In particular, each server must share information about the columns.Does match second column information from a server with the second column from another server? This is not an insurmontable problem and in fact is the fundamental problem of distributed data. There must be some coordination arrangements between servers. In this example, the server must be able to indicate - in a global-property corresponds to which column.

Table 8-1
School-leaving examination 2017, E written test, maternal language and literature, E writtnen test

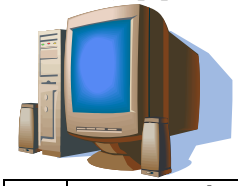
Tabel 8-1
Tabular data on students in a district that will support the school-leaving exam 2017, **E written test**

ID	Name student	E written test -subjects	Oral test result -A/B	High school graduation year
1.	Popescu Robert	The Romanian language	Level advanced	2017
2.	Ionescu Amalia	The Romanian language	Level average	2017

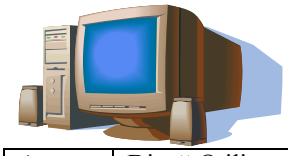
[1] Ph.D candidate, Polytechnics University, Bucharest, Romania, patrasculum@yahoo.com

3.	Poenaru Ciprian	The Romanian language	Level experienced	2017
4.	Dincă Otilia	The Romanian language	Level advanced	2017
5.	Schmitdt Ian	German	Level advanced	2017
6.	Toces Istvan	Hungarian	Level experienced	2017
7.	Ilie Damian	The Romanian language	Level experienced	2017
8.	Ivan Ludmila	Russian Language	Level advanced	2017

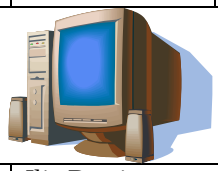
Figure 8-1 Distribution of Web data, row by row- adapted from Dean Allemang and Jim Hendler, [1] 2008



1.	Popescu Robert	The Romanian language	Level advanced	2017



4.	Dincă Otilia	The Romanian language	Level advanced	2017
6.	Toces Istvan	Hungarian	Level experienced	2017

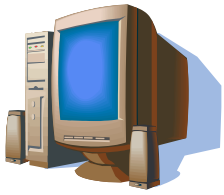


7.	Ilie Damian	The Romanian language	Level experienced	2017
3.	Poenaru Ciprian	The Romanian language	Level experienced	2017

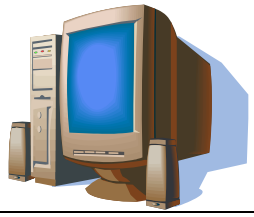
Figure 8-2 illustrates another strategy in which each server is responsible for one or more columns of the original table. In this example, one is responsible for the high school graduation year and the result of Oral test result -A/B and another on behalf of students. This solution is flexible in a way different from the solution in Figure 8-1. The solution in Figure 8-2 allows each machine to be responsible for a certain type of information.

Figure 8-2

Distribution of data on Web columns by columns adapted from Dean Allemang and Jim Hendler, [1], 2008



High school graduation year	Oral test result -A/B
2017	Level advanced
2017	Level average
2017	Level experienced
2017	Level advanced
2017	Level advanced
2017	Level experienced
2017	Level experienced
2017	Level advanced



E written test -subjects	Name student
The Romanian language	Popescu Robert
The Romanian language	Ionescu Amalia
The Romanian language	Poenaru Ciprian
The Romanian language	Dincă Otilia
German	Schmitdt Ian

If we are not interested in high school graduation year for each student need not take into account information on that server. If you want to specify something new about the students (student absent or present) can add a new server with information not to disturb the others.

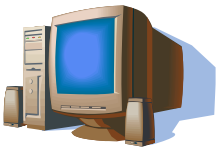
This solution is similar to the solution in Figure 8-1 in which it asks some coordination between the servers. In this case, coordination should relate to the identity of entities that must be described. How do I know if a server line 3 refers to the same entity row 3 of another server? This solution requires a global identifier describing entities.

The strategy outlined in Figure 8-3 is a combination of the two previous strategies, the information is not distributed any rows or columns instead is distributed cell by cell. Each computer is responsible for a number of cells in the table. This system combines the flexibility of the two previous strategies. Two servers can share a single entity description (in the figure, high school graduation year and name of the student are stored separately) and they can share the use of private property (in Figure 8-3. The oral test results for ID 6

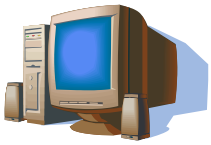
and 7 are represented on different servers). Flexibility is required if we want our system to support distributed data truly AAA slogan.

Figure 8 -3

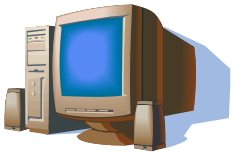
web data distribution, cell by cell -adapted from Dean Allemang and Jim Hendler,[1], 2008



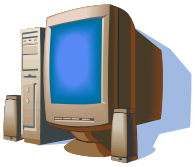
	Oral test result -A/B
ID 7	Level experienced



	Name student
ID 2	Ionescu Amalia



	E written test -subjects
ID 4	The Romanian language



	High school graduation year
ID 2	2017

	Oral test result -A/B
ID 6	Level experienced

Applying AAA slogan, every server needs to be able to make statements of any entity (as the case Figure 8-2) but any server also needs to be able to specify any property of the entity (as in the case Figure 8-1). The solution shown in Figure 8-3 has the advantages of both solutions.

The solution – distribution of data using RDF [1].

But this solution also combines the two strategies and costs. Not only that now we need a global reference for the column header but we need a global reference lines. In fact, each cell must be represented by three values: a global reference line, a reference column and value in the global cell itself.

This third strategy is called RDF

Because a cell is represented by three values, the basic building blocks are called RDF triple. ID for all calls are subject to triple. ID column is called the predicate of the triple (given that those entities properties specific columns by rows). Cell value is called the triple object. Table 8-2 presents triple in Figure 8-3 as a subject, predicate and object.

Triples become more interesting when more than one triple reffers the same entity that relates in Table 8-3. When more than one triple means the same thing, sometimes it is convenient to look at triple the direct graph where each triple is an edge of the object or subject, the predicate that the level of the edge.

Tabel 8-2 **Triple samples.**

Tabel 8-2 Triple samples.		
Subject	**Predicate** [2]	**Object** [2]
ID 7	**Oral test result -A/B**	Level experienced
ID2	**Name student**	Ionescu Amalia
ID2	**High school graduation year**	2017
ID 4	**E written test -subjects**	The Romanian language
ID 6	**Oral test result -A/B**	Level experienced

Tabel 8-3 **Triple samples.**

Tabel 8-3 Triple samples.		
Subject	**Predicate** [2]	**Object** [2]
Ionescu Amalia	High school graduation year	2017
Ionescu Amalia	Oral test result -A/B	Level average
Tokes Istvan	E written test -subjects	Hungarian
Tokes Istvan	Oral test result -A/B	Level experienced
Poenaru Ciprian	E written test -subjects	The Romanian language
Poenaru Ciprian	High school graduation year	2017
The Romanian language	E written test -subjects	Ionescu Amalia
German	E written test -subjects	Schmidt Ian

Conclusions

We described RDF as [9]"a way of distributing data across several sources". But when we want to use the data we need to combine these sources again. A triple representation of value is the ease with which this type of joint can be achieved. Since the information is represented as a simple triple combined information from two graphs is as simple as all triple graph formation of each individual graph, taken together. Therefore, representation of the data distributed on the Web in RDF is an optimal solution for collecting data to analyze Baccalaureate and provide feedback to the schools.

Bibliography

[1]Dean Allemang and Jim Hendler- Semantic Web for the Working Ontologist, Effective Modeling in RDFS and OWL, Morgan Kaufmann Publication,2008
[2] Legea nr. 1/5 ian 2011 a educaţiei naţionale
[3]The article "Intro to structOntology"(http://www.mkbergman.com/959/intro-to-structontology/
[4]Thomas R. Gruber – „A Translation Approach to Portable Ontology Specifications", Knowledge Acquisition, 5(2): 199-220, 1993, Knowledge Systems Laboratory, computer science departament, Standford University, Stanford, California.
[5]Matthew. Horridge - A practical guide to bulding OWL ontologies using Protégé 4 and]\O-ODE Tools, edition 3.1, The University of Manchester, March 2011
[6]Ioan Andone – Ontologiile şi modelarea informaţională a întreprinderii – Analele Ştiinţifice al Universităţii „Ioan Cuza", Iaşi, 2005-2006
[7]Natalya Noy and Deborah L. McGuines –„ Ontology development 101- a guide to creating your first ontology", Stanford University, Standford,2001
[8]Carola Eschenbach and Michael Gruninger - „Formal Ontology in Information Systems", Proceeding of Fifth Conference, FOIS, IOS Press, 2008.
[9]Gregoris Antoniou and Frank van Harmelen, - A Semantic Web Primer, second edition, The Mit Press, Cambridge, Massachusetts London, England, 2008