ONTOLOGICAL ENGINEERING FOR CORPORATE KNOWLEDGE PORTAL DESIGN[i]

Tatiana Gavrilova
Saint-Petersburg State Polytechnical University
Intelligent Computer Technologies Dpt.
195251, Politechnicheskaya 29/9, St. Petersburg, RUSSIA
Tel: + 07 812 371 82 72, Fax: + 07 812 373 06 89 gavr@fn.csa.ru
Vladimir Gorovoy
Intelligent System Lab., Institute for High Performance Computing and Data Bases, St.Petersburg, RUSSIA, E-mail: vgorovoy@mail.ru

The paper presents one approach aimed at corporate knowledge portal design based on the principles of ontological engineering. This approach proposes knowledge acquisition technique that scaffolds the process of information structuring by visual knowledge/data mapping as a powerful mindtool.
Ontologies that define the main concepts of the company's activity are used as a roadmap for knowledge management. The discussed technology is taught at the executive courses offered on the foundations of knowledge engineering and management to IT managers of different companies in Russia, Poland and Estonia. The described ideas are implemented in a series of computer programs that provide research/design framework for professional knowledge engineers and managers.

1. INTRODUCTION

New information age has forced large companies to have intranets with numerous hyperlinked pages. To create process information in such ill-structured representation media is not an easy task. Knowledge management (KM) – both as a new branch of management and as an information technology – is targeted to scaffold the information maintenance problems.

But the research focus in modern KM systems is still aimed at technology and software development aspects. While knowledge elicitation and structuring are really crucial and essential part of any non-trivial KM systems design, e.g. corporate portals. It is unfortunate that KM doesn't make into practice the experience gained in knowledge engineering and intelligent systems' development.

Traditionally conceptual domain structuring, which is the backbone for any corporate system, was often made by pen-and-pencil technique. Nowadays the situation has changed dramatically. The speed of information systems development and knowledge acquisition (KA) has drastically improved due to a range of novel technologies. During the last 10 years the main interest of researchers in this field was concerned with the special tools that helped in knowledge capture and

L. M. Camarinha-Matos et al. (eds.), *Processes and Foundations for Virtual Organizations*

structuring (Adeli, 1994; Wielinga, Schreiber, Breuker, 1992). These tools help knowledge engineers and experts to specify knowledge rules, check consistency, debug consultations and even suggest new rules (Adeli, 1994; Eisenstadt, Domingue, Rajan, Motta, 1990; Neches et al, 1991).

Many KA tools and methodologies assisted to cut down the lifecycle time required to refine, structure and test human knowledge and expertise (Wielinga, Shreiber & Breuker, 1992). The highlights in this area are related to early works in 80-ies and visual knowledge engineering (Aussenac-Gilles, 1993; Eisenstadt et al., 1990) is the next milestone which would aid knowledge acquisition. Such approach helps to traverse and organize visually an emerging knowledge store in rather natural form, for example as an «image panel» or a sketchpad for the visualization maps and diagrams. This would greatly benefit domain expert and users in making the maximizing the utility of the system. Such graphical techniques are called conceptual graphs (Sowa, 1986), concept maps (Lambiotte et al., 1984, Jonassen, 1998), visual knowledge representation languages (Kremer, 1998) or ontologies (Gruber, 1993).

2. KNOWLEDGE STRUCTURING

Although the popular methods described above are genuinely powerful and versatile, the knowledge analyst working for a real portal design is still weakly supported at the most important and critical stage in the knowledge engineering life cycle - transition from elicitation to conceptualisation. Understanding and realisation of the company information structure and expert's reasoning way makes this transition smooth.

A special methodology namely, the Object-Structured Analysis (OSA) has been developed (Gavrilova and Voinov, 1992-2000). OSA combines object-oriented approach with classical structured process-oriented methodology. OSA is intended to help knowledge engineer to perform from the most informal step of knowledge acquisition to concluding in informal conceptual structuring of company information. This approach enhances the classical structured analysis methodology (Yourdon, 1990) by supplementing with knowledge engineering methods.

OSA is based on decomposition of subject domain into (at least) eight strata (Tab.1). How ever the knowledge analyst is free to define the appropriate number of stratum. This methodological approach provides an effectual transition framework for KM systems.

s1	***WHAT FOR*** Knowledge	*Strategic* Analysis: Targets Aims, Requirements, and Constraints.
s2	***WHO*** Knowledge	*Organisational* Analysis: Developers Team, Human Resources, and Actors.
s3	***WHAT*** Knowledge	*Conceptual* Analysis: Main Concepts, Processes, Entities and Relationships between them.
s4	**HOW TO** Knowledge	*Functional Analysis*: Business Processes Modelling, Decision-Making Models.

s5	*WHERE* Knowledge	*Spatial* Analysis: Environment, Communications, etc.
s6	*WHEN* Knowledge	*Temporal* Analysis: Schedules, Time Constraints, etc.
s7	*WHY* Knowledge	*Causal* Analysis: Explanations to Decision-Making Models.
s8	*HOW MUCH* Knowledge	*Economical* Analysis: Resources, Losses, Incomes, Revenues, SWAT, etc.

Table 1 - Matrix for OSA

Level → / **Stratum↓**	**Domain Level (u_1)**	**Problem Level (u_2)**	**Sub-Problem Level (u_3)**	**...**	**(u_n)**
Strategic Analysis s_1	E_{11}	E_{21}	E_{31}	E_{i1}	E_{n1}
Organisational Analysis s_2	E_{21}				
Conceptual Analysis s_3	E_{31}				
Functional Analysis s_4	E_{41}				
Spatial Analysis s_5	E_{51}				
Temporal Analysis s_6	E_{61}				
Causal Analysis s_7	E_{71}				
Economical Analysis s_8	E_{81}				
...				E_{ij}	
s_m	E_{m1}				E_{mn}

Filling the matrix includes two steps:

Step1. *Global* (vertical) analysis results in process oriented decomposition of the heterogeneous company information into various levels of abstractions.

Step 2. *Local* (horizontal) analysis results in segregation of information into various *stratums,* each of which captures a property/ feature oriented structure.

The number of levels depends heavily on the peculiarities of the company and could vary dramatically for different strata. The OSA methodology recommends that the minimum number of levels at 3. A violation would indicate that the domain knowledge continues to be ill-structured.

The first level (or column 2 in the table) corresponds to the company information in general. The second one corresponds to the specific problem to be solved. The others may correspond with particular sub-problems, depending on the required reasonable deepness of detailing.

The minimal obligatory set of strata for the corporate portal development is:

s3: Conceptual Structure or company ontology (Gruber, 1993).

s4: Functional Structure or problem solving procedures and business models.

Other strata are developed if needed by system requirements, e.g. spatial and temporal analysis strata (s5 and s6) may be designed in the companies where the issues of scheduling, real-time operations and object manipulation are substantial.

Pseudo algorithm for Step 1 may be sketched in such form:

1.1: Gather all the data and knowledge about company processes identification and all other data that is revealed by prior knowledge elicitation (interviews, verbal protocols, observations, etc.)

1.2: Select a set of N strata to be formed (N≥3).

1.3: For each i-th stratum select a subset of all available information, relevant to that stratum and represent it in way appropriate to that stratum (see below).

1.4: If there remains unused bulk of information, increase the number of strata and repeat step 1.3. Otherwise, begin the horizontal analysis of each declared stratum.

Step 2 is horizontal analysis of strata that depends on the number of columns in OSA matrix and may be performed in two ways: deductive (top-down) and/or inductive (bottom-up).

The most essential stratum is s3 (WHAT-analysis) and resulting conceptual structure present a set of the domain ontologies.

The pseudo algorithm for step 2 procedure may be shown using s3 example:

2.1. Gather all the information relevant to WHAT-stratum, select and verbalise all essential concepts.

2.2. Reveal hierarchies among these concepts and represent them visually.

2.3. Detail the concepts via top-down strategy and form meta-concepts via bottom-up strategy. Exclude repetitions, synonyms, excessiveness and contradictions.

2.6. Discuss the concepts that were not included into that structure with the expert and either transfer them into other strata or exclude them at all.

2.7. Divide resulting graph into levels and represent accordingly to Tab.1.

Analogous algorithms were developed and practically tested and evaluated by the authors during several projects on corporate portal engineering.

3. VISUAL ONTOLOGICAL ENGINEERING

The basic philosophical definition and its further development (Stamper, 2000) are pointing that term ontology stands for study of "being". But in information science now ontology is a set of distinctions, explicitly made in order to understand and view the world. There is some of variety of definitions of this milestone term (Neches et al, 1991; Gruber, 1993; Guarino et al, 1995; Gomez-Peres, 1999):

1. Ontology defines the basic terms and relations comprising the structured vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary.
2. Ontology is an explicit specification of a conceptualisation or a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base.

On one hand all these definitions clarify the importance of ontological approach to knowledge structuring, and on the other hand gives enough freedom to the open thinking.

Ontologies may give an intuitively clear representation of company structure, staff, products and relationship among them. Ontologies are useful structuring tools, which could prove as an organising axis along which every employee can mentally mark his location in the information hyperspace of company. Sometimes expressing all information in a single ontology may be virtually impossible. This forces many corporate portals to include a set of ontologies. This in turn creates new problems, which arise when jumping from one ontological space to another. The web as it stands today offers no warning that the axis has shifted. It suffers from being a heterogeneous information space.

A necessary criterion for a knowledge engineer's software tool, which allows structuring of the subject domain based on the above-mentioned stages, should necessarily follow the phenomenological nature of the knowledge elicitation and are described in the above-mentioned algorithms. The tool must not frustrate the knowledge engineer with any hidden "game rules". Ideally, the tool must be self adapt to particular cognitive features of the knowledge engineer.

Moreover, each of the stage of analysis described above may be represented visually in its appropriate terms, an approach which is adopted in some commercial expert system shells.

CAKE-2 (Computer-Aided Knowledge Engineering) is a unique visual tool developed with the above-mentioned objective. It was initially developed by Mr. Alex Voinov and later enhanced by Mr. Tim Geleverya. CAKE illustrates the idea of knowledge mappability, a concept that finds application in the data mining and structuring of heterogeneous data base design.

CAKE-2 proposes a kind of a *visual knowledge representation language*, analogues of the same may be found in a group of visual ontology construction tools – ranging from PROTEGE-2000 (Noy, 2000) to Visio.

As an example, let us consider the stage s3 of the formation of conceptual structure of the domain. The example domain is taken from the practical work on corporate portal design for large audit and consulting company in St. Petersburg.

The ontology editor enables one to create new concepts representing the original knowledge source and relationship between them. Some of these concepts (regarded as "terminal") may have a set of "attributes or concepts" of very concrete nature (e.g. IT department and Training department on that picture). These attributes are listed below the concepts. The static relations (or "arcs") are drawn using either the drag-and-drop interface or through a menu-driven command box. In the described version of the system only a pre-defined set of link types is implemented (IS-A, HAS_PART, HAS_PROPERTY, IF, FUNCTIONS_AS). All the manipulation primitives of this part of the system are similar to modern picture editors (like Visio), except that CAKE knows more about the "nature" of its graphic objects.

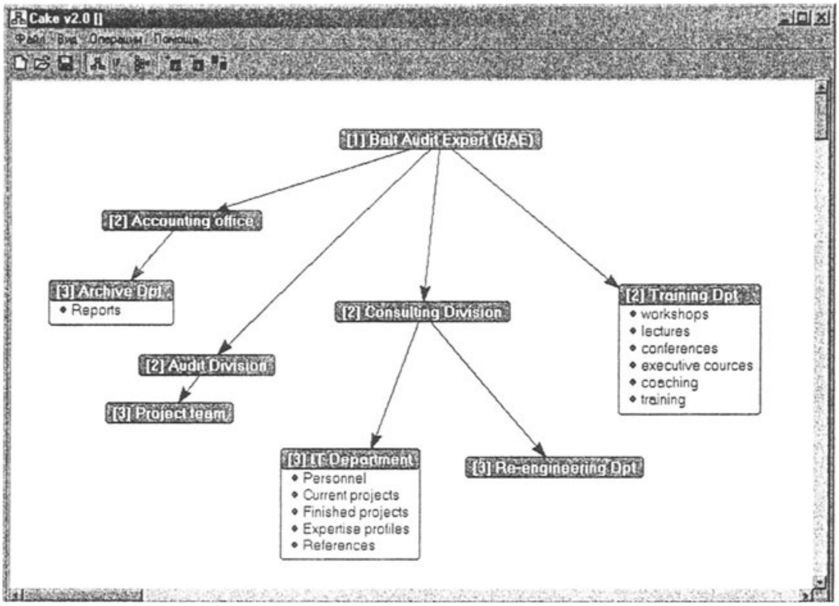

Figure 1 - an example of the domain structure

PORTO (PORTal Ontology) tool was developed to help build corporate web-portals. The main idea of the system is to provide a visual environment for knowledge engineers in intuitive conceptualisation of web-portals. The ontology developed in PORTO is problem-oriented. It consists of concepts of three types: container, element, and link. Concept of the container type can hold concepts of container and element types. Element type are terminals, it may contain concepts of the link type. Link type is used to encapsulate real links involved the designing web-portal.

The ontology developed in PORTO for the example described above (see fig.1) is shown on the fig.2. Main container, representing the start page, may be specified (bold font is used to show it). Files may be attached to concepts of container and element types to represent their contents. When the main container is defined and files are attached, a prototype of the designed web-portal can be made.

The described technology works as a mind tool for portal analyst and as a site-map for the users. It was implemented during last 2 years for development of different corporate information systems.

4. OTHER APPLICATIONS OF VISUAL SPECIFICATION APPROACH

The other applications of CAKE technology could help to implement knowledge sharing methods into company management, e.g. it may change the entire design cycle of the hypertext tutorials and database development. It forces the designer to define explicitly his knowledge and to follow the top-down technology versus the bottom-up one. The least but not the last contribution of the CAKE technology into this scope of problems concludes in the possibility for the end user to navigate *consciously* through the hypermedia space.

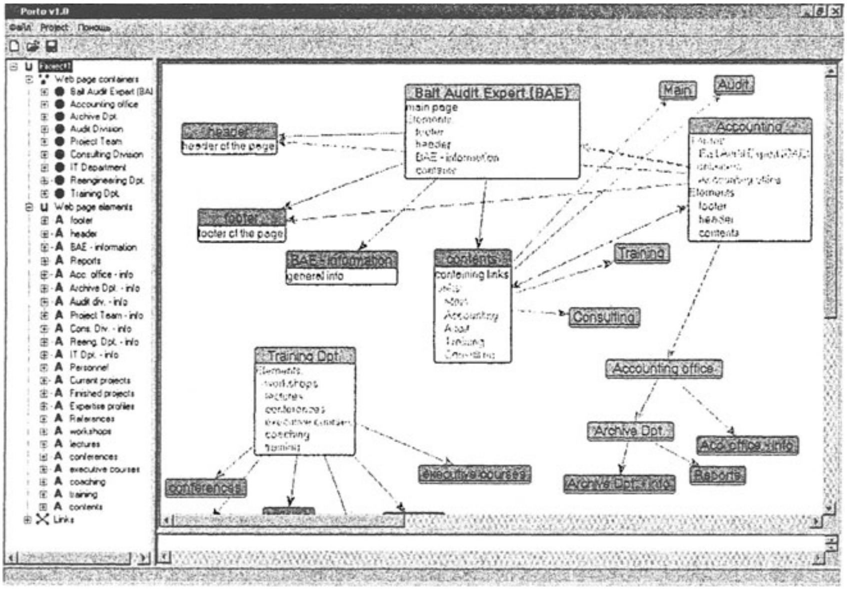

Figure 2 – PORTO layout

The OSA approach may be applied also to the hypertext design using heterogeneous knowledge sources. The term Heterogeneous *Knowledge Sources* is supposed to include: experts' and specialists' interviews, minutes, business communications, audio speech records, remote textual database queries and traditional textual information (reports, books, articles, newspapers, etc.) either in a hard copy or in an electronic form.

A better apprehension of each of such streams of data might be achieved by creating knowledge structure. Therefore, a visual editor for knowledge structure plays the role of a two-dimensional, pictorial conspectus of the regarded piece of information.

The imposing of the knowledge structure on such amorphous hyperlink spaces can dramatically shorten the conceptual apprehension of the corresponding flow of information. In this way, the CAKE technology, even in the described implementation, appears to be useful in this scope of problems, because it offers key functionality for elucidating of the basic logical skeleton of the domain. Even the plain visualising of the logical schemata of the domain has a powerful cognitive impact both on the user and designer.

5. DISCUSSION

Using visual paradigm for representing and supporting the knowledge acquisition process helps knowledge engineer to concentrate on the problem rather than on its details.

Ontological engineering based on tools like CAKE and PORTO may serve as semantic roadmap in company information space. And portal designers can also use it as a tool for structuring (intelligent reviewing) of the data incoming from *heterogeneous knowledge sources.*

Acknowledgements
The authors are very grateful to Alex Voinov and Tim Geleverya with whom they discussed the main ideas, and to Mathiew Koshy who helped in editing.

6. REFERENCES

1. Adeli H. *Knowledge Engineering.* McGraw-Hill. New-York, 1994.
2. Eisenstadt M., Domingue J., Rajan T., Motta E. Visual Knowledge Engineering, *IEEE Transactions on Software Engineering*, 1990, Vol.16, No.10, pp.1164-1177.
3. Gavrilova T. Computer-Aided Knowledge Engineering. Proceedings of Int.conf.*"Design methodologies in Microelectronics"* BENEFIT Concerted Action in the frame of the COPERNICUS programme // Vienna Industrial Day.-Vienna, 1995, pp.411-414.
4. Gavrilova T.A., Voinov A. Visualized Conceptual Structuring for Heterogeneous Knowledge Acquisition // Proceedings of Int. Conf. on *Educational Multimedia and Hypermedia* EDMEDIA'96, MIT, Boston, USA, 1996, pp.258-264.
5. Gomez- Perez A. Tutorial on Ontological Engineering: IJCAI'99.
6. Gruber, T. A translation Approach to portable ontology specifications. *Knowledge Acquisition*, 1993, Vol. 5. 199- 220.
7. Guarino, N.; Giaretta, P. Ontologies and Knowledge Bases: Towards a Terminological Clarification. In *Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing.* IOS Press. 1995. 25- 32.
8. Jonassen, D.H. Designing constructivist learning environments. In C.M. Reigeluth (Ed.), *Instructional design models and strategies*, 2nd ed. Mahwah, NJ: Lawrence Eribaum, 1998.
9. Kremer R. Visual Languages for Knowledge Representation, in Proceedings of *KAW'98 Eleventh Workshop on Knowledge Acquisition, Modelling and Management,* Banff, Alberta, Canada, 1998, pp.124-129.
10. Neches et al. Enabling Technology for Knowledge Sharing. AI Magazine . Winter. 36- 56, 1991
11. Noy N.F., Fergerson R.W., Musen M.A. The knowledge model of Protege-2000: Combining interoperability and flexibility. 2nd International Conference on Knowledge Engineering and Knowledge Management (EKAW'2000), Juan-les-Pins, France, 2000.
12. Sowa, J. F. Conceptual Structures: Information Processing in Mind and Machine. Reading, Massachusetts, Addison-Wesley, 1984.
13. Scott A., Clayton J.E., Gibson E.L. A practical Guide to Knowledge Acquisition, Addison-Wesley Publ, 1994.
14. Stamper R. New directions for System Analysis and Design. In J.Filipe (Ed.), Enterprise Information Systems, Kluwer Academic Publishers, Dordrecht, 2000.
15. B. Swartout; R. Patil; k. Knight; T. Russ Toward Distributed Use of Large- Scale Ontologies Ontological Engineering. AAAI- 97 Spring Symposium Series 138- 148.
16. Tu S.W, Eriksson H., Gennari J., Shahar Y. & M. A. Musen. Ontology-Based Configuration of Problem-Solving Methods and Generation of Knowledge-Acquisition Tools: Application of PROTEGE-II to Protocol-Based Decision Support. Artificial Intelligence in Medicine, 7(1995): pp. 257-289.
17. Wielinga B., Schreiber G., Breuker J. A Modelling Approach to Knowledge Engineering. Knowledge Acquisition, 4 (1). Special Issue, 1992.
18. Yourdon, E. Modern Structured Analysis. Prentice-Hall Bristol, 1989.

---

[i] The research was partially supported by Russian Foundation for Basic Research (grant 01-01-00224).