

Embedded Systems Knowledge Base

Klaus-Michael KOCH & Monia LOGE

Technikon Research, Villach, Austria
(koch@technikon.com, loge@technikon.com)

Wolfgang PREE & Sebastian FISCHMEISTER

University of Salzburg, Austria
(pree@SoftwareResearch.net, fischmeister@SoftwareResearch.net)

Abstract: Embedded systems play an increasingly important role in today's society. The spectrum of embedded systems ranges, for example, from mobile phones to X-by-wire control systems in cars and airplanes. The regional embedded systems knowledge base, called EmBase, is a comprehensive project which investigates and maps the knowledgebase in embedded systems with a focus on Austria and Central Europe. EmBase supports individuals, private and public institutions and companies that implement, develop, support, purchase or sell embedded systems so that they can successfully identify key market trends and directions and align themselves and their activities with these trends

Keywords:

Embedded systems, knowledge base, identification of know-how, who-is-who in certain field, representation of technology oriented news, Industrial project, Frameworks for knowledge portals, Management of distributed knowledge, Knowledge management in business process management

Categories:

A.1, A.m

1 Introduction

Embedded systems are playing an increasingly important role and have become a catalyst for change in the sectors of computing, data communications, telecommunications, industrial control, automotive, medical equipment and entertainment.

Currently, several transitions in the application domain of embedded systems are taking place:

The complexity of the functions to be covered is growing continuously.

Due to Moore's Law, cheaper and more powerful computing resources continue to support a shift from hardware-based to software-based systems.

So far, embedded systems have mainly been stand-alone solutions. Now, embedded systems are increasingly connected in networks, which needs more sophisticated design techniques incorporating communication technologies and security mechanism.

This shift in requirements will have a strong impact on the whole community involved in embedded systems, similar to the early stage of the introduction of the personal computer.

The EmBase provides structured information in the embedded systems field for users and potential users of such systems, points out general activities and knowledge, looks for relevant research activities, investigates the various products, identifies the companies and major players in the field, looks for ongoing co-operations, quantifies the need for personnel in the field of embedded system, and further qualification needs. Consequently, EmBase promotes regional networks of embedded systems activities.

Besides that, the EmBase Web site offers a partner search facility, e.g., for EU projects, a technologically oriented discussion panel, and a press-release section for promoting international and regional embedded systems activities.

Overall, the project provides access and creates awareness on embedded systems and promotes extensive deployment of high-potential products and solutions by the industry and research groups.

This paper will focus on the data model and the data acquisition as this project is still running. More information and results will be presented in consequent papers once the project is completed.

The primary goals, of the knowledge base project, are to provide access and a comprehensive overview of available technologies, competencies, companies, research institutions, education and training facilities, qualification structures and networks on embedded systems and their current use by the addressed parties in the observed region. The project shows global trends and visualizes the current as well as the future impact of the global movement in Austria and the surrounding region. Focus regions are Germany (Bavaria), Italy, Czech Republic, Slovenia and Hungary.

Furthermore the EmBase identifies possibilities and overcomes problems and shortcomings in all aspects regarding the education and training, qualification, information networks, etc. within the target group. Locating and seeding potential synergies will be made to overcome the identified deficits and strengthen future participation in European projects to come (e.g., Integrated Projects, STREP projects, Medea+, Leonardo, etc.). Finally the EmBase will map the need for personnel in the field of embedded systems.

The target group for the EmBase are current and potential new embedded systems technology users, development community, in present as well as future fields of applications. Suppliers and vendors of embedded system technology is another target group and also all organisations and people involved in research, education, training and consultancy in the area of embedded systems. Public organisations and funding agencies are the last but not the least important target group for the project.

2 Applied methods and technology

The scope of this EmBase project is to make a thorough analysis of the status quo of embedded systems in Austria and its surroundings with a focus on the most prosperous areas. In the first step, inquiries (web search; personal contacts, interviews) are used to generate a comprehensive list of companies active in embedded systems (development, design, production, etc., current activities (development, research, education), envisaged activities and activities in European research programmes.

The collected data is investigated in workshops for the three fields: software, hardware and education and are finally cross-analysed to provide a comprehensive information base. The knowledge of “how-to solve the problem” is actually the manager’s ability, knowledge and skill [2].

In the next consequent step, the data and analysis are used to build a web based knowledge platform, the EmBase-Web. The EmBase-Web becomes accessible in the second quarter of 2003 through www.embase.org. The key technology to set up and maintain the website is the one applied for the fully automated news editorial like the formerly published GoogleNews technology [1].

The challenge of EmBase is on the one hand to identify relevant sources for the area of embedded systems and to cope with dynamically changing addresses. In the case of EmBase, these sources are newspapers, TV stations, and scientific articles. The project, on the other hand, provides an excellent categorization (analogously to the categories Top Stories, World, U.S., Business, etc. in case of news) for embedded systems topics.

2.1 Entities in the observed region

In the project start-up the team made a crude first estimation of the number of organisations and companies involved in embedded systems in Austria. This orientation should help the consortium to define a sensible strategy for the EmBase project approach. The number of companies were estimated by partners in each of their field in order to plan for the data model, questionnaire planning, amount of workload and to get a feeling of suitable milestones and deliverables for the project. The number of estimated companies are shown below.

~ 500.000 (Companies in Austria)	Total	p < 200	p > 200
Total Number of Companies (AU)	2650	2345	305
Hardware oriented	450	405	45
Software oriented	500	420	80
System Integrators	400	380	20
Vendors	700	680	20
Research focus	200	165	35
Part Producers (Chip, PCP Board..)	400	385	15
Private Research Inst. (AU)			
Hardware oriented	25	15	10
Software oriented	35	23	12
Higher Education Centers (AU)			
Hardware oriented	14	4	10
Software oriented	30	5	25
Research focus	40	30	10
Number of publ. Research Inst. (AU)			
Hardware oriented	30	23	8
Software oriented	50	35	15

Table 1 Number of Entities (*p* stands for number of full-time employees)

Before data acquisition started, the consortium sorted out what embedded systems and Fit-IT means to them as the area of embedded system is highly complex with several parameters that have to be taken in account. Due to the current shift of requirements for embedded systems, the question is also what is not an embedded system.

2.2 Core competence Screening

An early phase of the EmBase project is the screening of the core competences of the observed entities in the observed region. The outcomes of this phase will serve as a basis for the creation of a data model for the knowledge base, and provide an overview of the companies, organisations research facilities etc thus facilitating further data acquisition.

This phase of the project consist of gaining access to the existing profiles for the observed entities, review the profiles to gain an overview of the entities, review the Intranet data available for the entities and gathering organisation such a cluster networking web sites and research the relevant expert literature to determine the appropriate method of screening the core competences. In this core competence phase a concept for the screening process will be created which specifies the data required from the companies to identify their core competences and the required tools which are needed to enable the data capture.

In order to gain information from the companies and institutes, computer based questionnaires will be sent out. In the questionnaire free space should be left for

possible (new) ideas of the companies/institutes. It can not be ruled out that companies/institutes may in some areas have more information than we propose to be filled out. By doing this the EmBase project can also help companies/institutions to find out more about themselves.

The resulting first crude data set uses the Microsoft vcard presentation. That should include enough information for further decisions to be taken on the data set. Interviews are also performed on the entities to get a more thorough information base for a core competence analysis. .

3 Data Analysis and Model

The goal of the project is to gain accessible knowledge about the embedded system Scenery in Austria and the surrounding countries and to gain a comprehensive overview and a manageable set of data. An extremely important part of this knowledge management project is to determine what information is required from the entities and how this information is linked to each other. The goal of this phase is to identify the information necessary to build a knowledge management map of the entities. The information is described by means of a data model. The goal of this phase is to define the rules for an appropriate data model to be used to acquire information from the companies.

Another factor that needs to be taken into account is repeatability, the data acquisition must be ongoing so that the knowledge management map can always provide an up-to-date picture of the entities. The data-model and the data linkage need to be sustainable enough even for future technology trends. This aspect will be considered in more detail in phase of the data maintenance.

The first step in this phase is analysing of the core competence data obtained and evaluate the method of data acquisition employed. Consultation with knowledge management experts and science and research advisors are made in order to determine suitable types of data model and methodology and to identify research relevant information. Thereafter the first draft of the draft of the data model is created which is based on the questionnaire. An important task within this phase is to consider how to make the data model maintainable and sustainable and what effect the selected knowledge management tools will have on the data model.

It is also important to acknowledge that a key to the communication and use of existing, and creation of new knowledge, lies in the transformation of implicit knowledge to explicit structures perceivable und usable by others when acquisitioning, analysing of data and making the model.

The aim of this project phase is to select the appropriate knowledge management tool for the embedded knowledge base. The selection will be based on the draft data model and analysis. In this phase specification are set up for the criteria for the knowledge base tool and evaluation of which tools are appropriate for the EmBase project and available within the allocated budget and timescales. Thereafter a tool will be selected and tested for its suitability and finally evaluated and implemented.

Our approach will start from the identification of companies' needs, relating them to the most adequate concepts, methodologies and tools for exploring knowledge assets, e-work and virtual teams, as well as research and training for information and knowledge exchange. The consortium will diagnose user problems, organisational and technological approaches for managing critical knowledge assets and development of strategic solutions using existing innovating tools. EmBase will also adapt core competence and change management concepts to assessing knowledge management needs in the virtual network formation process.

On the basis of the results of the analysed situation the research team will define network mission, synergy and value creation potential, anticipated life cycle, integration level and value creation as drivers for virtual networking and introduction of knowledge management practices.

The knowledge base is seen as an evolving platform which will serve and support the entity community far into the future. This phase aims to determine a strategy for the long term maintenance of the knowledge base. Aspects that need to be addressed by this strategy are how to keep the data up to date, e.g. repeatability of data acquisition, how to purge out-dated data from the platform, how to maintain the knowledge base, e.g. secure access, performance and efficiency, ongoing analysis and finally how to migrate tool and data in response to changing requirements in future.

4 The Software Platform

As the EmBase knowledge platform includes 100+ links, data acquisition and categorizing needs to be done automatically – therefore realized within a software framework. Frameworks are in effect skeletal or generic applications which can be fleshed out and customized into full-fledged software programs [3]. As the data sources and the data itself may change rapidly, the software platform must be well structured and modular.

4.1 Platform Overview

The software platform can be roughly divided into three separate modules/components:

- **Retriever:** responsible for fetching the data from the data sources and converting it into a common format.
- **Analyzer:** responsible for mining the data and identifying hot topics.
- **Publisher:** responsible for generating news from the results of the analyzer and distributing it.

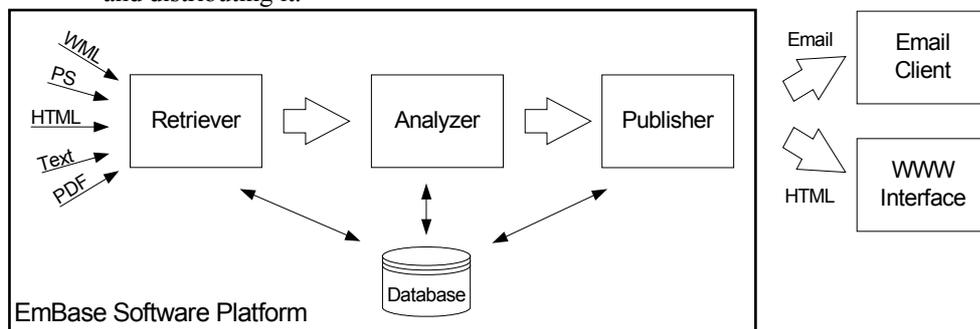


Figure 1. Overview of the EmBase Software Platform.

4.1.1 The Retriever Component

The retriever component has a list of data sources. This list contains URLs to companies, educational and research facilities. The retriever fetches the data published at this URL and converts it into a common format, which can be processed by the analyzer component.

To make the latest news visible, the retriever scans through this list once a day. Thereby, it has to identify the URLs that contain new data. This is partly done by comparing the file date and header data. If the source data has changed, the retriever will fetch eventually new data.

After fetching the data from the source, the retriever has to convert the existing data into a common format. The retriever can process the following formats: HTML, WML, plain text, Postscript, PDF, Microsoft WinWord. Furthermore, it can process the following zipped formats: TAR, Gzip, and Zip. After the retriever component identified the file type, it converts it to plain text (it removes all figures and other

fancy extras) and puts it in the database. In addition to the plain text, the retriever component also stores the creation date, the copyright information, the link to the source data, the source name, and metadata in XML (e.g., is it a new item, what was its source format).

4.1.2 The Analyzer Component

The analyzer component retrieves the plain-text data from the database, prepared by the retriever component. The job of the analyzer component is, to mine the data and find correlations between articles and topics. Once such a link is found, it marks the data for publishing. For example, four data sources post articles related to drive-by-wire. The analyzer component will find all three articles, see that they talk about the same topic, and decides whether this makes it a “hot” topic or whether this is just a lucky coincident.

To find such articles, the analyzer component categorizes and prioritizes the data. It splits the texts and categorizes them according to given keywords. Every keyword has a priority, which is used for sorting the articles. Several keywords together make a category. This list of keywords and the corresponding categories are also stored in the database and can be adapted and expanded dynamically.

Based on the relevance of the keywords and the number of occurrences within the article, the analyzer component determines the importance and category of the article. Then, the analyzer component will group articles that cover the same topic. If it identifies a critical number of articles that cover the same topic, it will mark the articles as important and the topic will be marked as “hot”.

4.1.3 The Publisher Component

The publisher component receives the data from the analyzer component. It extracts all articles marked as relevant and also all topics marked as “hot”. The publisher component has two output channels: web-pages and electronic mail.

The publisher component reads the database and searches for articles marked as relevant. It then creates news postings for each article. Depending on the copyright of the article, the publisher component will either only publish the title and the initial 100 characters of the article and a link to the source, or it will publish the whole article.

Published news are distributed through web-pages and electronic mail. The publisher component uploads all web-pages to the EmBase portal, through which the public can access it. Furthermore, the publisher component distributes the news postings to registered users via email.

4.2 Deployment

The EmBase software platform is spread across four nodes. The design rationales are the amount of processing time required to accomplish the task, task latencies (e.g., a network timeout for sending an electronic mail), and specialized nodes (e.g., the portal node and the email distribution node are a pure web-server and a pure email server, respectively).

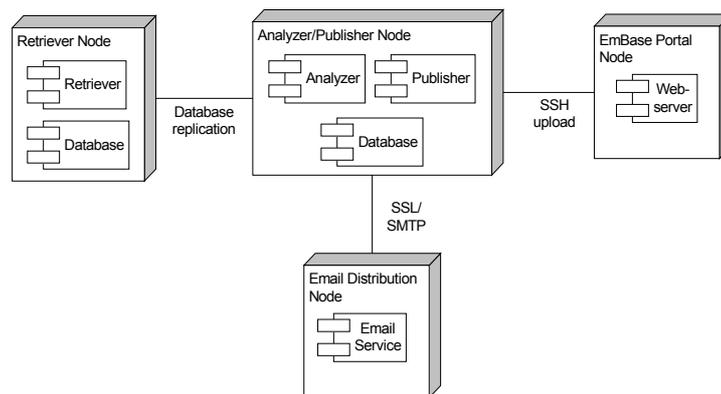


Figure 2. Deployment Diagram. [object Management Group]

The retriever node hosts the retriever component; it uses the database locally. Once the retriever finished its job, the database is replicated to the Analyzer/Publisher node. This node hosts the analyzer and the publisher component. When the database has been replicated, the analyzer component starts mining the data. After it finished, the publisher component creates the news postings. The publisher component also distributes the news postings to the EmBase portal node and the email distribution node. The web pages are uploaded to the portal node using the secure shell upload mechanism. Once the pages are uploaded they are immediately available to the public. The email postings are distributed by the email-distribution node. The publisher component communicates with the email-distribution node via a secure SMTP connection. Once the email-distribution node received all messages, its mail transfer agent delivers the electronic mail to the recipients.

5 Conclusion

A multidisciplinary team of researches are building up an embedded systems knowledge base. The approach to derive a well suited data model in systematic way for the knowledge base is presented. The software developed for fetching the data from the data sources, for identifying hot topics and generating news and distributing high-lights it is shown.

Acknowledgment

The EmBase project is supported by the FIT-IT embedded system initiative which is a new Austrian research support programme that concentrates on high-quality research in the area of information and communication technologies. We gratefully thank the Austrian federal ministry of transport, innovation, and technology for it throughout support.

References

- [1] Google News, <http://news.google.com/> 2400 Bayshore Parkway, Mountain View, CA (USA), 94043 USA
- [2] Nonaka I., Takeuchi H.: "The Knowledge-Creating Company."; University Press, New York; 1995
- [3] Mohamed E., Fayad.; Douglas, C. Schmidt.; Ralph E. Johnson.: "Building Application Frameworks" : Object-Oriented Foundations of Framework Design, John Wiley & Sons, 1999