Workflow Technology Supporting Information Filtering From The Internet

Witold Abramowicz and Jacek Szymanski

Department of Computer Science at University of Economics in Poznan, Poland, {w.abramowicz, j.szymanski}@kie.ae.poznan.pl

ABSTRACT

We live in the information era. Companies need only relevant information to perform. Therefore a problem of selecting information from information noise arises. There can be some processes identified that make up the task of selecting information. Such processes without optimisation and use of specific technologies can be very time and resource consuming. This paper will discuss the problems of obtaining relevant information and automation and simplification of those processes in business perspective.

INTRODUCTION

In the recent years new market trends have been observed. New technologies are introduced almost every day. Companies need to act very quickly on rapidly changing markets. Therefore the need of information arises. Companies require information about their competition, market situation and technologies to survive on their markets.

Having to little information is no longer an issue. Now we have to select pieces of information from the excess of information we are surrounded with. Therefore companies need a technology to help them to react fast and accurately to changing environment.

The information needs of companies are quite constant. They concentrate on certain fields and do not change much in time. Therefore a technology called Information Filtering can be used in those companies in order to improve their information performance.

FILTERING

Information filtering assumes that there is a competent user who wants to be submitted with information on a certain subject. The subject from the point of view of the user does not change or changes little.

The problem is how information can be automatically processed and how to specify the information needs of the user.

In the Internet information is contained in documents. Documents contain messages, which are collections of signals in a given language. [Panel 94] Messages themselves carry no information. They need to be accompanied with context in order to give information. The same message can bring different information depending on context. Therefore only messages with context carry information. [Sveiby 98]Because in the Internet messages are contained in documents, to analyse information, one needs to analyse documents.

The analysis of documents consists on comparing them to users' information needs. The needs can be specified by a set of keywords (terms) describing the fields of interest of users. By relevant information we understand information that is in a way connected to the given subject and is interesting to the user from the point of view of the subject. The set of terms forms a simple structure (that can be further developed) called profile. Users' profiles describe the users' information needs. The terms can be paired with weights describing their importance for the subject. In profiles terms are also paired with their desired frequency in documents. Profiles of users can be combined to create a multi-profile. This is done to improve the performance of filtering systems. Multi-profiles are not just a "sum of profiles". The general rule is that if a document is relevant to the multi-profile it is relevant to at least one profile. If it is not relevant to the multi-profile it is not relevant to any profile.

Documents are compared to the user profiles by creating certain statistics helping developing a mark. A document profile is created. Then the profile is compared with users' profiles. With this comparison one can compute certain measures to describe the document's relevance (in this case similarity) to the profile. For their profiles users can specify the minimum value for one or more of each of available measures for documents to be evaluated as relevant. [Balt] [Belkin Croft 92].

The whole process is managed by filtering systems, which are capable of obtaining and analysing documents. There are two ways new documents can be obtained: passive and active. In the passive mode, the system just waits for documents to be sent to it from their sources. This method is usually used with payable information sources. The other way of obtaining documents is active searching for documents. The use of agents' communicativity, reactivity and pro-activity is very helpful here. [Foltz 92]

After having read a document user has to score it. The document can be scored either by the user or by the system basing on observation of the user while reading the document. Both ways have disadvantages and have to be chosen basing on the capabilities of the system.

If a document has been found highly relevant by the user it means that the terms used in this document more likely present the information needs of the user. Therefore the system can apply some changes to the user's profile by changing terms and their weights. Thanks to this mechanism the profile is improved and documents submitted to the user is submitted by more relevant documents.

Relevant documents can then be processed in order to use them in knowledge bases. With such solution not only employees' personal knowledge is broadened but also corporate knowledge of a company can be improved.

WORKFLOW

A filtering system for only one user would be a small one but one could think of using such systems in companies. But when the system grows to serve more and more users the processes in it become more and more difficult to manage effectively. Therefore there is a need to improve performance of such systems.

Workflow technology can be a solution to this problem. Workflow concentrates on automating and simplifying processes in which humans take part. Although the research in the field of Workflow has been done for a number of years now, it is still relatively new to many parts of business activity. Companies seem not to realise possible benefits from introducing this technology.

Contrary to what some think, Workflow assumes human interaction in the processes. The processes are managed by Workflow System. The structure of Workflow is however much deeper than the level of processes. [WfMC RM]

Workflow System stores information about models of processes. Those models describe a theoretical order and dependencies of elements of processes.

Real business life processes are then instances of those models. Analogy to class-object relation is helpful here. Each process in Workflow is an instance of a model.

Processes consist of a set of activities. An activity can be perceived as a single step in the process. It can be manual or automated and involve human and/or machine resources. Workflow however concentrates on activities that can be automated. As well as processes activities are also instantiated, which means that real representation of theoretical models appear. [WfMC TG]

Each activity is performed by a Workflow Participant. [WfMC TG] The notion of Workflow Participant usually applies to human resources but it can also be applied to machine resources. Intelligent software agents are an example.

Models of simple processes are usually linear. But as the complexity of processes grows new structured have to be introduced in order to describe reality. Processes can contain subprocesses, they can also be performed in parallel. Workflows can be subject to decision results and can iterate an activity or a block of activities. Therefore there is a possibility of executing the same activity several times until an exit condition is met. [WfMC TG]

From the above description one can easily see that Workflow terminology and concepts allow modelling and thus creating processes with variety of structures. The abundance of elements is needed in order to enable modelling even the most complicated parts of business reality. Workflow technology can be introduced in different departments of companies from production to marketing and therefore improve more processes within them and cross-department processes.

WORKFLOW SUPPORTING INFORMATION FILTERING

Since Workflow simplifies and automates processes one could wonder if it could not be applied to information filtering tasks. It could improve certain parts of filtering tasks by automating and managing it. [WA]

Usually in companies numerous user profiles are created in order to reflect different information needs of each user. With each new profile the complexity of the system and the amount of calculations needed grows. In order to simplify this part, a multi-profile can be created.

The process of profile improvement can be managed by Workflow System by conducting the order of tasks required to change profiles. What is more, Workflow System can monitor the changes and inform the filtering system administrator about any changes that can influence the performance of the system. Workflow System can force users to score documents which simplifies the problem of document scoring.

The documents that were found relevant can be further processed in order to include them in the company's knowledge base. Therefore a need for creating knowledge base user profiles arises. Those profiles can be managed similarly to the filtering system profiles and therefore also managed by Workflow System.

The use of Workflow System in companies goes beyond filtering tasks. When introduced to supporting filtering of information from the Internet, Workflow System can also be used to manage other processes. Combining filtering with HR processes is an example. [WAFS]

From the technological point of view it might be interesting for companies to introduce Workflow technology to manage intelligent software agents used to gathering information from the Internet. Searching for documents, monitoring document sources and co-operation between agents can be managed by Workflow System. Better performance, less irrelevant documents and lower risk of obtaining bad materials are the possible benefits.

SUMMARY

In the era of information companies require technologies that would help them obtain only relevant information. Information filtering can be a solution here. With implementation of user profiles and a multi-profile for the company, the employees of the company are provided only with relevant information. However, managing filtering systems require human interaction. Therefore a way of simplifying and automating information-filtering processes is required.

Workflow technology can be of help here. Because it is possible to model almost any part of business reality for Workflow, it is possible to user Workflow not only for supporting information filtering but also for other parts of business activity.

Workflow technology can support information filtering in a variety of ways:
- automation and support of intelligent software agents for gathering information;
- simplification and management of profile improvement tasks;
- co-ordination of profile and multi-profile management;
- use of filtered documents for corporate knowledge bases;
- management of new profiles, new users.

Although both technologies described here have been known for some years now, the combination of both is still a matter of future.

REFERENCES

[WA] Witold Abramowicz, *Information Filters Supplying Management Information Systems, WebNet 2001, World Conference on the WWW and Internet,* Orlando, Florida, October, 24-27, 2001, invited speaker

[WAFS] Witold Abramowicz, Filip Stankowski, *Instancing Workflows with Information Filtered from Internet,* A. J. Baborski, R. F. Bonner, M. L. Owoc (eds)., Knowledge Acquisition and Distributed Learning in Resolving Managerial Issues, Mälardalen University Press, 2001, 145-155, ISBN 91-88834-22-0

[Balt] *Information Filtering And Retrieval: Overview, Issues And Direction; Basis for a Panel Discussion* (Moderator: Nicholas DeClaris)

[Belkin Croft 92] Belkin Nicolas J., Croft Bruce W. *Information filtering and information retrieval: Two sides of the same coin?*, 1992

[Foltz 92] Foltz Peter W. *Personalized Information Delivery: An Analysis of Information Filtering Methods* http://www-psych.nmsu.edu/~pfoltz/cacm/cacm.html

[Lehnert b] Lehnert Wendy *Information Extraction.* http://www-nlp.cs.umass.edu/nlpie.html

[Marinilli] Marinilli M., Micarelli A., Sciarrone F. *A Case-Based Approach to Adaptive Information Filtering for the WWW* http://wwwis.win.tue.nl/asum99/marinilli/marinilli.html

[Panel 94] *Shannon Theory: Present And Future - Panel Discussion* http://it.ucsd.edu/IT/Newsletter/94_december/2shannon.html

[SAIC] *Science Applications International Corporation WebSite* www.muc.saic.com/main.html

[Sveiby 98] Sveiby Karl-Erik *What is information?* www.sveiby.com.au/information.htm

[Takkinen 96] Takkinen Juha *Information Retrieval and Information Filtering (IRIF), Spring 1996: Introduction to Course*, 1996

[WfMC RM] (David Hollingsworth) *Workflow Management Coalition Reference Model*, 1995

[WfMC TG] *Workflow Management Coalition Terminology and Glossary*, 1999

[Wiener 48] Wiener Norbert *Cybernetics*, MIT Technology Press, 1948

[Zaenen Uszkoreit] Zaenen Annie, Uszkoreit Hans *Language Analysis And Understanding*

Related Content