Information Systems in Management
Primary version of the journal is an electronic version.

Editor
Department of Informatics, Warsaw University of Life Sciences – SGGW

Editorial Committee
Dr hab. inż. Arkadiusz Orłowski – Editor-in-Chief
Dr Piotr Łukasiewicz – Executive Editor
Dr inż. Tomasz Żąbkowski – Scientific Editor
Prof. nadzw. Kris Gaj – Linguistic Editor
Dr hab. Wiesław Szczaśny – Statistical Editor

Editorial Council
Dr Oguz Akpolat – Mugla Sitki Kocman University, Turkey
Prof. dr hab. inż. Ryszard Budziński – Uniwersytet Szczeciński
Prof. dr hab. Witold Chmielarz – Uniwersytet Warszawski
Dr hab. inż. Leszek Chmielewski – Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
Prof. Jeretta Horn Nord – Oklahoma State University, USA
Prof. Frederick G. Kohun – Robert Morris University, USA
Prof. Yuiry Kondratenko – Black Sea State University, Ukraina
Prof. Alex Koohang – Middle Georgia State College, USA
Prof. Vassilis Kostoglou – Alexander Technological Educational Institute of Thessaloniki, Greece
Prof. dr hab. Marian Niedźwiecki – Uniwersytet Łódzki
Dr hab. inż. Arkadiusz Orłowski – Szkoła Główna Gospodarstwa Wiejskiego w Warszawie – Chairman
Dr hab. inż. Joanna Paliszkiewicz – Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
Prof. Kongkiti Phusavat – Kasetsart University Bangkok, Thailand
Prof. Josu Takala – University of Vaasa, Finland
Dr hab. inż. Ludwik Wicki – Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
Prof. dr hab. inż. Antoni Wilński – Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

Address of the Editor
Faculty of Applied Informatics and Mathematics, WULS – SGGW
ul. Nowoursynowska 166, 02-787 Warszawa, Poland
e-mail: isim@sggw.pl, www.ism.wzim.sggw.pl

ISSN: 2084-5537
e-ISSN: 2544-1728

Wydawnictwo SGGW
ul. Nowoursynowska 166, 02-787 Warszawa, Poland
e-mail: wydawnictwo@sggw.pl, www.wydawnictwosggw.pl

Table of contents

Grzegorz Arkit, Silva Robak, Aleksandra Arkit
APPLYING E-LEARNING SYSTEMS FOR BIG DATA EDUCATION ....................... 85

Krzysztof Berbeka, Sandjai Bhulai, Ewa Magiera
DECISION SUPPORT SYSTEM FOR WATER ADAPTING PRICING POLICY ........ 97

Waldemar Karwowski, Joanna Sosnowska, Marian Rusek
THE RECOMMENDATION ALGORITHM FOR AN ONLINE ART GALLERY ...... 108

Gustaw Konopacki
ASSESSMENT OF THE DEGREE OF FITTING THE TRANSPORT POTENTIAL
OF THE TRANSPORTATION COMPANY TO A RANDOM DEMAND
FOR TRANSPORT SERVICES ................................................................. 120

Tomasz Rokicki
E-COMMERCE MARKET IN EUROPE IN B2C .......................................... 133

Tomasz Woźniakowski, Piotr Jałowiecki, Krzysztof Zmarłowski,
Magdalena Nowakowska
ERP SYSTEMS AND WAREHOUSE MANAGEMENT BY WMS ....................... 141
APPLYING E-LEARNING SYSTEMS FOR BIG DATA EDUCATION

Grzegorz Arkit a), Silva Robak b), Aleksandra Arkit b)

a) Faculty of Computer Science, Electrical Engineering and Automation
   University of Zielona Góra (UZ)

b) Faculty of Mathematics, Computer Science and Econometrics
   University of Zielona Góra (UZ)

Processing massive data amounts and Big Data became nowadays one of the most significant problems in computer science. The difficulties with education on this field arise, the appropriate teaching methods and tools are needed. The processing of vast amounts of data arriving quickly requires the choice and arrangement of extended hardware platforms.

In the paper we will show an approach for teaching students in Big Data and also the choice and arrangement of an appropriate programming platform for Big Data laboratories. Usage of an e-learning platform Moodle, a dedicated platform for teaching, could allow the teaching staff and students an improved contact with by enhancing mutually communication possibilities. We will show the preparation of Hadoop platform tools and Big Data cluster based on Cloudera and Ambari. The both solutions together could enable to cope with the problems in education of students in the field of Big Data.

Keywords: Big Data, e-learning platform, Hadoop platform tools, cloud computing, Linux, virtualization
1. Introduction

One of present-day big challenges in information systems are the issues associated with coping with and utilization of the vast amounts of data and big data [23]. There are some definitions and terms concerning Big Data usage. Unfortunately, because this is a new subject, there is no one strict (official) definition of these terms.

**Massive (or large) dataset**, in the simplest way, is a dataset, which cannot be stored on single computer: usually this is a dataset with size at least a few TB (there is no strict limit). Such a dataset may contain many physical files; it may be homogeneous (e.g. datasets from Large Hadron Collider [11]) or may contain data in various formats (e.g. all kind of documents in an enterprise).

**Big Data** is more than massive (large) dataset – in this case we do not consider only a size of dataset, there are other important features. One definition uses 3V’s characteristic [3] (it covers mainly technical parameters):

1. **Volume** – the size of stored data;
2. **Velocity** – how fast new data is generated and how fast we can access this data;
3. **Variety** – the type and nature of data (structured vs. unstructured, different kind of files: text, spreadsheets, music, pictures, movies and so on).

At present one may consider also other characteristics like [2,20,21]:
4. **Veracity** – messiness or trustworthiness of data; in some cases data is worthless if it’s not accurate;
5. **Value** – how we can turn our data into value;
6. **Variability** – may describes few aspects, e.g.: number of inconsistencies in the data, changing speed at which big data is loaded into your database;
7. **Visualization** – how to present data in visual form to see some dependencies;
8. Others: **Validity, Vulnerability, Volatility, Viscosity, Virality**, etc..

In this paper we use the first 3V. Moreover, one of most important feature of Big Data (compared to RDBMs or data warehouses) is a different approach of data processing: for RDBMs (and data warehouses) data should be strictly structured before loading to database – as for Big Data, the data is stored in raw form, then the transformations are done later by the target system.

For treatment of big data we should have dedicated infrastructure (with appropriate hardware and software tools) to store and process it. Such an infrastructure should be scalable and fault tolerant.

The another important aspects associated with Big Data are data analysis methods and stakeholders cooperating in Big Data projects. In our paper we will consider only a second aspect that is stakeholder, especially how they may collaborate in teams in projects. It is presented in Fig. 1.
1.2. Goals of the work

As mentioned in above Section 1, the presented subject is relatively new, the difficulties with education on this field arise. There are two main aspects of the problem.

The first is a choice of an software platform for processing Big Data during student courses. Needed is a platform, which is free and regarded as a quasi-standard, with many tools available for diverse kinds of tasks and also accompanied with additional study materials (books, web pages, forums, etc.). In the paper we will show that such platform is Apache Hadoop [8]. By limited time resources we should show an appropriate subset of Hadoop tools.

The second aspect is the way of teaching in a course student course, especially with regard of group collaboration in a project. Big Data is a vast knowledge domain, difficult to comprehend by individuals collaborating in teams. We do not aim, specialists strictly concerned with one selected field, but prefer omnibuses over specialists.

Another related dilemma related to the first aspect is how to run a Big Data platform on regular personal computer (PC). The production systems for Big Data are based on extended hardware platforms, but our available hardware resources are limited.
The aspects mentioned above lead us to two main goals of this work:

1. How to prepare a hardware platform for teaching Big Data courses, which can be launched on typical lab computer? A software solution should be enriched by additionally data samples and some prepared tasks and examples. Moreover, a hardware solution (based on free software) should be easy to launch on a student’s computer. In the following Sections 2 and 4 we will consider application of some constituents of a Hadoop platform.

2. How to improve the teaching paths and a knowledge transfer process and at the same time collaboration in teams? We will focus on this topic in Section 3 and propose a usage of an e-learning tool.

In Section 5 we conclude our work.

2. Hadoop – a scalable software platform for distributed computing

An important question is what kind of hardware and software infrastructure should be used for processing large sets of data? It is obvious that it would be too much for single computer, we also know that e.g. RDBMs systems [6] can be scaled only up to a fixed limit. So, we need a system which may be scaled linear, and with a reasonable costs. We need to increase computing power by adding computers instead of replacing them.

There is such a solution - Hadoop which is a scalable software platform for distributed computing. Hadoop can store practically unlimited size of data and can process this data in distributed environments. It is an open-source, free solution, and relatively simple to scale-up. Of course we must keep in mind, that a scaling hardware generates additional costs.

Hadoop platform contains three important modules:

- HDFS – Hadoop Distributed File System;
- YARN – a framework for job scheduling and cluster resource management;
- MapReduce – a YARN-based system for parallel processing of large data sets.

In a distributed Hadoop environment each part of data is stored in several copies (usually at least 3) on different computers (i.e. cluster nodes). One of the most important assumption is that data is processed locally. It means that data is processed where it is stored (on the same computer/node) which minimizes the network transfers. Moreover, the system is fault tolerant, i.e., when one of the nodes fails, the results from this node are lost. So only a repetition of the calculations on from a broken node are needed. What is more, such situations are managed by the system itself, so user do not need to undertake additional actions.

Now we will give short description of main Hadoop platform items: HDFS and MapReduce.
2.1. **HDFS – Hadoop Distributed File System**

As mentioned in the beginning of this Section, in *Hadoop* each file stored in a file system is divided into several parts (blocks), and each part is stored in several copies on different locations. In Fig. 2, a schema of processing model in *HDFS* architecture is presented. It is worth mentioning, that *HDFS* structure is *rack-aware*, which means that if you have a cluster build with nodes in many racks, its data will be distributed so as to minimize the effects of the failure of the whole single rack.

Commands used for managing files in *HDFS* are very similar to those used in *Linux* operating system [1, 5, 7].

![Diagram of HDFS architecture](image)

**Figure 2.** Processing schema in *HDFS* architecture

2.2. **MapReduce model**

*MapReduce* is a framework (with tools and methods) for parallel processing data in *Hadoop* environment. There are:

- a *map* operation in which for every record we calculate the key-value pairs; all pairs with the same key will be in the same group, e.g. if we process weather data and for every observations we calculate a pair year-temperature, all observations with the same year will be in the same group;
- a *reduce* operation in which for each group we calculate some features from value (aggregate), e.g. for every group of observations calculated in previous point, we may calculate for instance a maximum temperature.

This scheme is very similar to grouping and aggregating information in SQL [22].

89
We may create an appropriate map and reduce procedures (methods) in many programming languages (i.e. JAVA or Python), or we may use more specialized tools, like Pig [1, 5] – a tool with an own language similar to SQL.

2.3. Hadoop – choosing tools for teaching courses

We mentioned above only three tools connected with Hadoop: HDFS, MapReduce and Pig. But still, there are much more such tools available. We should keep in mind, that we should provide tools for creating complete analytical applications (paths) for solving given Big Data problems (see Fig. 3.).

There are many diverse tools for solving different problems (see Fig. 3), so we had to choose which tools should we use for our aims (there is no time available for teaching too many tools in one course). We have chosen the following applications: HDFS (Hadoop Distributed File System), Avro (data serialization system), MapReduce (system for parallel processing of data) and Pig (language and system which simplify MapReduce processing).

In Section 4 we will continue with further information on preparing Hadoop platform for a Big Data student course.

3. Moodle – an e-learning software platform

Even if we choose a limited set of big data tools, a range of learning materials is still large. So we must use tools dedicated for improve learning paths, like an e-learning platform to improve cooperation between a teacher and students, as well as directly between the students.

Such a tool should give us possibilities:
- to prepare and publish additional materials (tasks, exercises and examples, etc.), visible only for dedicated group(-s) of students;
• for certain tasks to request answers or solutions (within given time limits); we may observe them and react (comment or correct). Often tasks are multi-stage, and before going to the next stage students must complete the previous tasks;
• to report by the students some problems, doubts, propositions or solutions; it gives the teachers some opportunities to learn something new and/or improve a course;
• to prepare own materials and to share them with others students (and teachers);
• to present the obtained students’ results; this gives opportunities for verification and comparison of code, solutions, data, indirect results, etc., with other team members and to discuss.

Apart from the primary purposes (for the teachers), a cooperation should give the students opportunities to influence on a course and to extend knowledge according to their interest. There are opportunities to learn and improve team work.

For our Big Data course decided to chose Mood  e [4,10] - a free software tool for supporting traditional lecture and laboratory. It is a worldwide known e-learning platform, which can be used to enhance traditional forms of teaching by sharing links and various resources or making possible new additional interactions in communication (not limited to e-mail and personal contacts). This platform is very popular and appreciated in universities. It is used by several faculties in our university as well. Thanks to a wide range of functionalities like: sharing files, glossary, wiki, links, quizzes, forums, chats, blogs, workshops, Moodle fulfils our needs of improving cooperation between teachers and students, as well as directly between students. It also helps to stimulate students to be more active and responsible for their education. Moreover, very important to us is:
• multilingualism – usage of several languages simultaneously (configurable);
• availability of extensive documentation (including Polish language) - due to the great popularity of such software, there is a large community of users (also Polish), so it is easy to get a help;
• a mobile-compatible use interface and a cross-browser compatibility;
• a customizable interface - there are many ready-to-use themes (free of charge or paid), available on many web pages, including Moodle page [10].

For installing it from scratch (for training or testing purposes), one must have a valid web account with PHP handling, an access to SQL database (usually MySQL, especially for external web providers) and an e-mail box (for outgoing information). For installing Moodle the following actions are needed: download of a Moodle package (compressed archive) and unzipping it to the web account. Next one should start the own page in web browser and follow the instructions to complete installing Moodle.
4. Preparing *Hadoop* platform for a Big Data student course

It will be shown how to prepare *Hadoop* platform with specific tools on a single computer. Of course a relatively modern computer, capable enough to run typical tasks at home or laboratory (see below). Such a computer usually makes use of a *Windows* operating system, and this leads to the first problem because the tools for processing Big Data are based on *Linux*. The proposed solution is a virtualization platform – below will be shown how to use *Oracle VM VirtualBox* [13] for running a separate machine with *Linux* (Fig. 4.).

For preparing such a platform basic knowledge of using *Linux* (including installation) is required. In addition, we should have a computer with an operating system and *VirtualBox* installed. Such a computer should contain at least 8GB of memory (we need memory for a host operating system and additionally for a virtual machine). Computer with at least two cores is recommended. *VirtualBox* can replaced with any other virtualization platform (e.g.: *VM Ware* or *QEmu*) [14, 15].

![Figure 4. Preparing Hadoop platform on Virtual Machine VM](image)

Only the basic steps for creating a virtual machine with operating system *Linux* and *Hadoop* platform will be shown.

*Linux*

- prepare your favourite distribution of *Linux* (download ISO image): we use *Ubuntu* (with *XFCE* graphical desktop – recommended, but not required);
- create a virtual machine with at least 4GB of memory (RAM) and at least 200GB of disk space (HDD) – not all this space will be required for running machine (virtual disk will be expanded on demand); if you have computer with small disk you may use an external drive (USB 3.0);
- install *Linux* from ISO image: usually it is enough to confirm default options in an installer;
• install Guest Additions – some additional features (like fitting screen size of Linux to Windows operating system window size or ability to exchange files with the host system) will be given;
• install Java (Java Development Kit);
• recommended: add an ability to run commands with admin access rights without password to current user (sudo command in Ubuntu);
• recommended: install Midnight Commander (mc) – visual file manager.

Hadoop (HDFS, MapReduce)

In the simplest scenario only downloading of Hadoop binaries (version 2.7) and unpacking it to any directory is needed. Next, script variables HADOOP_HOME and JAVA_HOME in hadoop-env.sh file must be setup. In this way a working instance of your own Hadoop environment is made available. It is standalone version with no dedicated file system (HDFS) – file system for this version of Hadoop is common with base Linux file system. It is recommended to add a path to Hadoop binaries directory to your system PATH variable.

For using such system run hadoop-env.sh file should be executed at first. Then execute tasks (e.g. example WordCount included in Hadoop distribution).

If a system with separate (dedicated) HDFS system is desirable, few files should be modified [8]. Finally, for the first usage one should format HDFS (hadoop namenode –format) and execute services (start-all.sh). It should be considered, that Pseudo-Distributed Mode is a usage of one machine, that makes files (blocks) replication impossible.

For installation also additional tools will be used, but description of installing each of them is out of scope of this paper. Some of these tools require additional tasks to install them (i.e., compiling from source), so it is recommended to installing system tools like pip, git, snappy, ant, maven, and so on. The tools installed in our implementation are: Python environment, Avro, Pig and Eclipse IDE (Java Developers edition with plug-ins for editing Python and Pig scripts).

Preparing an own Hadoop cluster with additional tools

In the beginning of this Section we considered preparing Hadoop platform on a single computer. A very valuable experience is to show students both: the process of creating a cluster for Big Data and the benefits from using it. Below it will be shown how to prepare a Hadoop cluster.

First of all, to build a cluster, an appropriate hardware resources are required:
• NameNode – computer which will act like a monitor and supervise the operation performed in our cluster; we will use computer with 64GB of RAM;
• **DataNode** – computer which stores the files and processes them; we will use 6 computers with 24GB RAM (at least 3 to be able to show the principles of HDFS);

• all computers are created as a virtual machines in virtualized environment (Xen), each machine has allocated 4 processor cores.

Because any block in HDFS is stored on at least 3 machines (default), for each 1MB of data we need 3MB of disk space (plus space for operating system, tools and space for calculating/processing data). We should remember that when calculating the required disk space. Moreover, for the same reason, to see how HDFS system works (data partitioning and replication) we should use at least 4 machines (see Fig. 2.).

Our cluster is based on **Linux (Ubuntu, [9])**, so only two versions of the machines should be prepared: **NameNode** and **DataNode** (only operating system, prepared with the same procedure as for single machine – see Section 4). In virtualized environment it is possible to copy machines.

It is possible to prepare your **Hadoop** cluster directly from binaries, but in this case the configuration files need to be edited manually. A much simpler way is a usage of an integrated distribution like **Cloudera [16]** or **Ambari** [17], which contain additional integrated tools. For installing such distribution (we have chosen Cloudera) its binaries should be downloaded into your **NameNode** machine, executed (install Cloudera Manager) and one can launch a web browser with a proper link (local machine address with a specific port). From browser the machines on which you want to install the software (**DataNode**) can be selected, and the desired tools. Installing software separately on each machine is no not need.

**Hadoop cluster application example**

*Data size:* 900MB; weather data from NOAA [12], text files, unpacked, with minor corrections.

*Task:* calculate minimal and maximal temperature for each year (1901 – 2016).

*Results* (processing time):
• for a single node Pseudo-distributed mode (one computer): about 21 hours;
• for a presented cluster: 2 hours 13 minutes.

Other options for training Big Data are for instance:
• ready-to-use virtual machines with Big Data tools: **Oracle BigData Lite Virtual Machine, Hortonworks Sandbox on VM, Cloudera QuickStart VM** (required 4-8GB of RAM for machine);
• dedicated cloud solution (free of charge): **IBM Analytics Demo Cloud (Ambari)**: 4 machines with 32 cores each, 3x64GB RAM (**DataNode**) and 1x256GB RAM (**NameNode**);
• commercial clouds (but with free starting period or starting credit): *Amazon Web Services* [18], *Microsoft Azure* [19].

5. Conclusion

For teaching students in the domain Big Data appropriate tools and methods are needed. We have introduced an approach by using an e-learning platform *Moodle* and *Hadoop* platform tools (processed on the single machine and at the *Hadoop* cluster).

As for the experiences with the e-learning platform *Moodle* we can say that it met the initial expectations in concerning cooperation with the students. We can state, that after implementation, the effectiveness of teaching in our course as to knowledge sharing has increased, compared to the previous education cycles. Nevertheless, the much more additional time is required for the staff to prepare teaching materials available on the platform. Even so, there is still a low awareness of students in the field of cooperation within the groups.

Considering prepared *Hadoop* platform (for one computer) we can state that it can be started on any relatively modern computer. The process of creating such a machine can be carried out independently with open, free software. In our course the virtual machine consists basically of two files, easy to upload (both the first version and then the updates). A teacher can prepare the software on his computer and then upload or replace it. A limitation may be a file size – with large files and network 100Mbit speed the upload takes a few hours.

The snapshots feature allows to save the machine state and, after class, to restore the machine to its original state. This can also be done by copying the virtual machine file, but due to its size much more time is needed. The feature of machines’ separation allows each student a usage of its own machine.

Thus, the prepared solutions have greatly improved the education of in Big Data domain, the effectiveness of teaching has been increased. However low students’ awareness of the need for team collaboration needs improvements and this will be our next goal by using additional *Moodle* capabilities, such as wikis.

Our plans for the future also include usage of the extended the Big Data tools set with NoSQL databases and tools for data visualization and presentation.

REFERENCES


DECISION SUPPORT SYSTEM FOR WATER ADAPTING PRICING POLICY

KRZYSZTOF BERBEKA a), SANDJAI BHULAI b) EWA MAGIERA c)
a) Institute of Economics, Finance and Management, Jagiellonian University, Krakow, Poland
b) Department of Mathematics, Vrije Universiteit, Amsterdam, The Netherlands,
c) Institute of Computer Science, University of Silesia, Sosnowiec, Poland,

In the paper, the conception of Enterprise Information Portal (EIP) as an end-user interface of Simulation and Modeling System for Business (SMS-B) is presented. The system is a proposition of Business Intelligence education platform. EIP portals are also a base for Enterprise Integration Platform (EIP II) introduction in information and communication system in an institution.

Keywords: dynamic pricing system, water prices, efficiency of water consumption, demand management

1. Introduction

Nowadays, Integrated Support System for Efficient Water Usage and resources management (ISS-EWATUS) [1] is the outcome of the international research project entitled “Integrated Support System for Efficient Water Usage and resources management.” The ISS-EWATUS consist of four subsystems:
1. decision support system (DSS) for the efficient water usage at households,
2. DSS for efficient water management at municipal water company,
3. social-media platform: enabling and promoting water-saving behaviour, development and simulation of adaptive water price systems,
4. the adaptive water pricing system developed to assess the implications of current and optimal water pricing policies.

The ISS EWATUS adaptive pricing module is centred on research that consists of residential water consumption. The purpose of the module is to evaluate different pricing schemes to assess the viability of such schemes. In the paper, we provide an overview of the functionality of the dynamic pricing module. Moreover, we explain the different features of the module to correctly assess pricing schemes.

2. Validity of the adaptive pricing DSS

The dynamic pricing tool has been developed based on many steps that have been researched. First, water consumption data of both Poland and Greece has been analyzed to discover the impact of the tool. The unit water demand in Poland has been decreasing over the last 10-15 years as far as households connected to water supply systems are concerned. Analysis of municipal companies shows that current unit water demand in households fluctuates between 90 and 110 liters per inhabitant per day. Further decrease in households seems to be unfeasible because of the minimum water demand necessary to meet human needs. The influence of water prices, although difficult to be precisely assessed, does not seem to be significant. This renders the applicability and the arguments for applying adaptive water pricing void for Poland. For Greece, there is added value in the use of the tool. According to current practices, the water price is simply added up by the components of water supply cost, sewerage and waste treatment cost as well as infrastructure-improvement cost. Up to date, water policy pricing in almost all water utilities in Greece is not designed with any estimation, or prediction of the consumers’ reaction to a change imposed in the water price. So, demand elasticity is ignored completely!

Based on the findings of the research, an adaptive pricing model has been developed with its corresponding computing infrastructure. The parameters of the model are based on data analysis and the literature. Research conducted for the estimation of the drop in water demand caused by changes in water tariffs and especially demand price elasticity for water in Greece is indicated only in a few projects [2–5]. The research differs by size of the agglomeration, sample of the investigated inhabitants and the research period. The review was made with the differentiation of the calculation starting from average prices and marginal prices (the average price is defined as the water bill paid by the consumer divided by the amount of water consumed. The marginal price is the price that a consumer should pay, according to the water price structure, for the next m³ of water).

The impact of changes in the pricing policy by implementation in practice is a sensitive and delicate task that is influenced by many economic and social factors.
In practice, one tries to integrate on the one hand principles of free economy and on
the other hand social justice. In both Poland and Greece, one implements a low-
cost pricing policy while aiming at full cost recovery. In Poland, however, the wa-
ter demand level is very low, leaving little room for water reduction through adap-
tive pricing. Adaptive pricing in Greece has more added value, but due to the eco-
nomic crisis, there is a tension in affordability of pricing when full cost recovery
needs to be achieved. These reasons prohibit the adoption of adaptive pricing in
practice, limiting the validation of adaptive pricing to simulation studies (see also
recent research on this topic with the same conclusions [6]).

The simulation study was carried out by assessing multiple scenarios and their
impact. We conducted two major analysis. The first analysis was focused on using
the tariffs as a tool for demand management. This instrument, however, has quite
limited applicability and the associated consequences seem to be not affordable due
to political and social reasons. The reduction of water demand by 20% requires an
increase in the average price by 87% in the best scenario, and by 190% in more
realistic scenarios. The second analysis was geared towards an increase in the net
income of the water operator. An increase of 10% in the income, would require an
increase by 16% in the average price of water. This will typically hurt the most
sensitive of all clients.

3. The adaptive pricing module

The adaptive pricing module consists of four tabs with various statistics on
pricing schemes. The main page has a focus on global characteristics of pricing
policies. The input to the module consists of the price demand elasticity for a nor-
mal season (Q1/Q4) and the tourist season (Q2/Q3) (i.e., how does water consump-
tion by consumers change as the price changes), and the pricing policy (which con-
sists of the flat rate and the prices as a function of different water consumption
brackets). The user can input these quantities in the grey area on the main page of
the module, see Fig. 1 below. The prices are specified in euros per m$^3$, and the
brackets indicate the region in m$^3$/calculation period (quarter) in which the price is
in effect.

Once the parameters have been given by the user, the parameters can be sub-
mitted to the module by the submit button. The tool will then display the new pric-
ing policy together with the benchmark policy. The benchmark policy is the policy
that was in use a specific year to create a benchmark with the new pricing policy.
One can see an example of this graph in Fig. 2.
4. Statistics

Statistics of the new pricing policy are calculated automatically and are depicted in a table on the main page. The table includes the total water consumption in m³ over a year, as well as the breakdown into the different quarters of the year. It also lists the total revenue based on the water consumption not considering any effect of demand elasticity of consumers. This gives an intuitive idea of how much one would receive in revenues by changing the pricing schemes while everything else remains constant (this quantity is also specified per quarter). However, a change in the pricing schemes comes with changes in water consumption. Hence, the total revenue that one really collects is different. This is listed in the total revenue after the change in the water consumption, which has also a breakdown in quarters.

In the following example (see Fig. 3) one can see that a new pricing scheme is in effect. The scheme is such that if the water consumption would remain the same
(at the level of 51,949 m³ per year) for this population, then the total revenue would increase by 11.46% (from 68,216.10 euro to 76,034.28 euro). However, due to the fact that the water consumption decreases with 9.22% the total increase in revenues is actually less than 11.46%. It turns out that the revenues increase only by 0.27%. This example directly illustrates the difficulty in assessing pricing schemes. In cases where one expects that the revenues go up while at the same time water usage is reduced, it turns out that in practice different numbers correspond to reality. This warrants the development of a dynamic pricing tool.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Benchmark</th>
<th>Current</th>
<th>Rel. difference (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total water consumption (in m³)</td>
<td>51949.00</td>
<td>47198.20</td>
<td>-8.27</td>
</tr>
<tr>
<td>Total water consumption Q1 (in m³)</td>
<td>16545.05</td>
<td>14741.54</td>
<td>-10.15</td>
</tr>
<tr>
<td>Total water consumption Q2 (in m³)</td>
<td>13875.05</td>
<td>11771.08</td>
<td>-15.41</td>
</tr>
<tr>
<td>Total water consumption Q3 (in m³)</td>
<td>17366.00</td>
<td>14659.41</td>
<td>-15.57</td>
</tr>
<tr>
<td>Total water consumption Q4 (in m³)</td>
<td>10158.00</td>
<td>8195.14</td>
<td>-19.48</td>
</tr>
<tr>
<td>Total revenue before change in water consumption (in euro)</td>
<td>69216.10</td>
<td>76204.28</td>
<td>10.21</td>
</tr>
<tr>
<td>Total revenue Q1 (in euro)</td>
<td>12113.26</td>
<td>11418.37</td>
<td>6.46</td>
</tr>
<tr>
<td>Total revenue Q2 (in euro)</td>
<td>17176.33</td>
<td>16288.76</td>
<td>5.41</td>
</tr>
<tr>
<td>Total revenue Q3 (in euro)</td>
<td>26660.34</td>
<td>24897.31</td>
<td>6.71</td>
</tr>
<tr>
<td>Total revenue Q4 (in euro)</td>
<td>12837.15</td>
<td>11219.04</td>
<td>13.46</td>
</tr>
<tr>
<td>Total revenue after change in water consumption (in euro)</td>
<td>68216.10</td>
<td>68432.82</td>
<td>0.27</td>
</tr>
<tr>
<td>Total revenue Q1 (in euro)</td>
<td>15113.05</td>
<td>13045.36</td>
<td>-13.02</td>
</tr>
<tr>
<td>Total revenue Q2 (in euro)</td>
<td>17170.33</td>
<td>15701.11</td>
<td>-8.48</td>
</tr>
<tr>
<td>Total revenue Q3 (in euro)</td>
<td>29052.34</td>
<td>26447.82</td>
<td>9.19</td>
</tr>
<tr>
<td>Total revenue Q4 (in euro)</td>
<td>12837.15</td>
<td>12337.56</td>
<td>4.02</td>
</tr>
</tbody>
</table>

**Figure 3.** Statistics of a new pricing scheme

The tab with the details gives insight into the underlying data that the dynamic pricing tool works with. It is based on consumer data on water usage over a specific period. The table shows the customer code, the number of people in the household of the consumer, the water usage over the different quarters, the water bill in each quarter, the new price for the consumer of the new policy per quarter, and the water usage based on the new price per quarter. The columns BM Cons 1 to BM Cons 4 denote the water consumption in m³/household for that quarter in the baseline scenario. The columns BM WB Q1 to BM WB Q4 depict the expenditures in the baseline scenario in euros for that quarter. The columns WB Q1 to WB Q4 are the expenditures in the new scenario in euros for that quarter. Finally, Cons Q1 to Cons Q4 denote the water consumption in the new scenario in m³ per household. The data on this page essentially aggregates to the statistics on the main tab. The dynamic pricing tool internally calculates for each household the effect of price changes using the price demand elasticity and then aggregates this data on the statistics tab. Figure 4 gives an impression of part of the data on the detailed level.
Figure 4. Detailed information on the household level

<table>
<thead>
<tr>
<th>Customer Code</th>
<th>Household 1</th>
<th>BM Cons Q1</th>
<th>BM Cons Q2</th>
<th>BM Cons Q3</th>
<th>BM Cons Q4</th>
<th>BM WB Q1</th>
<th>BM WB Q2</th>
<th>BM WB Q3</th>
<th>BM WB Q4</th>
<th>WB Q1</th>
<th>WB Q2</th>
<th>WB Q3</th>
<th>WB Q4</th>
<th>Cons Q1</th>
<th>Cons Q2</th>
<th>Cons Q3</th>
<th>Cons Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001223456</td>
<td>5</td>
<td>-0.23</td>
<td>-1.33</td>
<td>-1.43</td>
<td>2.04</td>
<td>-7.33</td>
<td>-8.43</td>
<td>-9.54</td>
<td>-8.43</td>
<td>21.50</td>
<td>21.50</td>
<td>21.50</td>
<td>21.50</td>
<td>18.43</td>
<td>18.43</td>
<td>18.43</td>
<td>18.43</td>
</tr>
<tr>
<td>0001223457</td>
<td>2</td>
<td>-0.55</td>
<td>-1.66</td>
<td>-1.76</td>
<td>2.38</td>
<td>-8.43</td>
<td>-9.54</td>
<td>-8.43</td>
<td>-8.43</td>
<td>21.50</td>
<td>21.50</td>
<td>21.50</td>
<td>21.50</td>
<td>18.43</td>
<td>18.43</td>
<td>18.43</td>
<td>18.43</td>
</tr>
<tr>
<td>0001223458</td>
<td>4</td>
<td>2.34</td>
<td>2.46</td>
<td>3.45</td>
<td>4.56</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
</tr>
<tr>
<td>0001223460</td>
<td>5</td>
<td>2.34</td>
<td>2.46</td>
<td>3.45</td>
<td>4.56</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
</tr>
<tr>
<td>0001223461</td>
<td>4</td>
<td>2.34</td>
<td>2.46</td>
<td>3.45</td>
<td>4.56</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
</tr>
<tr>
<td>0001223463</td>
<td>5</td>
<td>2.34</td>
<td>2.46</td>
<td>3.45</td>
<td>4.56</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
</tr>
<tr>
<td>0001223464</td>
<td>4</td>
<td>2.34</td>
<td>2.46</td>
<td>3.45</td>
<td>4.56</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
</tr>
<tr>
<td>0001223466</td>
<td>5</td>
<td>2.34</td>
<td>2.46</td>
<td>3.45</td>
<td>4.56</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
</tr>
<tr>
<td>0001223467</td>
<td>4</td>
<td>2.34</td>
<td>2.46</td>
<td>3.45</td>
<td>4.56</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
<td>3.45</td>
</tr>
</tbody>
</table>
When a pricing scheme has been set the main page calculates several statistics, of which the most important ones are the effect on water consumption and total revenues. However, for a complete picture, it is necessary also to evaluate how this impacts the household on an individual level.

The details already give some insight into individual behaviour, however, the influence allows for further analysis. The table provides a sorted overview of the top 10 households that are affected most in several ways. One can select the top 10 households that have the highest expenditure, but also the top 10 households that have the highest reduction in water consumption. The table lists the customer code, the number of people in its household, the total water consumption and water bill under the benchmark policy, and the total water consumption and water bill under the new pricing policy. The last two columns display the difference in the expenditure of the consumer (in %) and the difference in water consumption (in %). Both of these columns can be sorted on to generate different top 10 listings. Fig. 5 depicts a screenshot of the table. One can see that the household set by this policy are affected such that there is an increase in expenditures of at most 5.59%. At the same time, this household, as a result of price changes, will use 14.44% less water.

There are many different pricing schemes that can be devised and evaluated. The different combinations of the flat rate, the volumetric brackets and the respective prices therein are immense. Therefore, one needs to be assisted in the evaluation of different pricing schemes. The simulation tab provided help in this requirement.

<table>
<thead>
<tr>
<th>Customer Code</th>
<th>Household</th>
<th>BM Consumption</th>
<th>BM Waterbill</th>
<th>Consumption</th>
<th>Waterbill</th>
<th>Rel. exp. diff. (in %)</th>
<th>Rel. cons. diff. (in %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1387000000</td>
<td>2</td>
<td>62.00</td>
<td>79.63</td>
<td>93.05</td>
<td>69.97</td>
<td>5.99</td>
<td>-14.64</td>
</tr>
<tr>
<td>1286000000</td>
<td>2</td>
<td>66.00</td>
<td>91.49</td>
<td>85.05</td>
<td>66.94</td>
<td>4.46</td>
<td>-15.46</td>
</tr>
<tr>
<td>1105000000</td>
<td>2</td>
<td>68.00</td>
<td>84.15</td>
<td>80.53</td>
<td>68.55</td>
<td>3.29</td>
<td>-12.27</td>
</tr>
<tr>
<td>1168000000</td>
<td>2</td>
<td>72.00</td>
<td>66.07</td>
<td>83.75</td>
<td>96.51</td>
<td>6.17</td>
<td>-11.02</td>
</tr>
<tr>
<td>1130000000</td>
<td>3</td>
<td>73.00</td>
<td>68.03</td>
<td>90.61</td>
<td>82.40</td>
<td>3.05</td>
<td>-10.91</td>
</tr>
<tr>
<td>1090100000</td>
<td>2</td>
<td>75.00</td>
<td>68.03</td>
<td>90.61</td>
<td>82.40</td>
<td>3.05</td>
<td>-10.91</td>
</tr>
<tr>
<td>1169000000</td>
<td>2</td>
<td>75.00</td>
<td>68.03</td>
<td>90.68</td>
<td>82.40</td>
<td>3.05</td>
<td>-10.93</td>
</tr>
<tr>
<td>1130200000</td>
<td>1</td>
<td>77.00</td>
<td>68.34</td>
<td>89.69</td>
<td>89.76</td>
<td>4.89</td>
<td>-10.48</td>
</tr>
<tr>
<td>1150300000</td>
<td>2</td>
<td>84.00</td>
<td>81.28</td>
<td>85.17</td>
<td>85.28</td>
<td>4.02</td>
<td>-12.23</td>
</tr>
<tr>
<td>1008200000</td>
<td>1</td>
<td>58.00</td>
<td>73.08</td>
<td>90.24</td>
<td>61.36</td>
<td>4.87</td>
<td>-13.36</td>
</tr>
</tbody>
</table>

**Figure 5.** Influences on the household level. Source: own preparation

This module simulates a large number of randomly generated pricing policies for a given price demand elasticity and flat rate. It changes the volumetric brackets and the prices therein. For each simulation, the module records several statistics, in particular, the difference in water consumption (in %) and the difference in revenues (in %) simulated over a period of one year. The module displays the Pareto frontier of the different pricing schemes, i.e., all combinations of the two performance indicators. Fig. 6 displays the Pareto frontier for a specific setting (in this case, an elas-
ticity of -0.4 and a flat rate of 12 euro). The results show that it is hard to obtain both an increase in revenues and reduction in water consumption at the same time.

One the one hand, an increase in prices has such an impact on the reduction of water consumption that it will not generate additional revenues. On the other hand, lowering pricing is also a possibility. This will increase the water consumption, but will not generate sufficient demand that additional revenues are generated. Hence, the curvature of the graph. Note that there a few policies that do attain a higher revenue while at the same time also reduce water consumption (there is a point at -6.5% in water consumption and 1.09% in revenues), however, the final result depends on the value of demand price elasticity index. These are quite rare though and indicate that setting a good pricing scheme is a difficult problem that needs to be approached with considerable care.

The price demand elasticity is a very important factor in this analysis. If the elasticity changes from the value −0.4 to −0.3, then there are more policies that attain better performance in water consumption and revenues simultaneously. In Fig. 7 we can see how the graph changes as the elasticity changes. It clearly shows that there are more points above the zero-line for the revenues. A similar analysis can be done with the flat rate.

![Figure 6. Pareto frontier of different pricing schemes](image)

![Figure 7. Pareto frontier of different pricing schemes with reduced price elasticity](image)
The graph depicts the realm of possibilities of different pricing schemes. In order to get more insights into the type of policies that obtain these results, a table is presented with all simulated policies and their performance.

<table>
<thead>
<tr>
<th>Policy (brackets, prices)</th>
<th>Diff. water cons.</th>
<th>Diff. revenues</th>
<th>Diff. final revenues</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.35, 130, 262, 292, 294, 206, 207), (0.075, 1.812, 1.946, 1.989, 1.992, 1.997, 2.2)</td>
<td>32.10</td>
<td>-17.72</td>
<td>7.81</td>
</tr>
<tr>
<td>(0.36, 43, 198, 280, 291, 298), (0.197, 1.501, 1.835, 1.852, 1.971, 1.974, 1.999, 1.998)</td>
<td>12.97</td>
<td>6.20</td>
<td>4.04</td>
</tr>
<tr>
<td>(0.30, 45, 117, 148, 248, 277, 294), (0.298, 0.651, 1.552, 1.867, 1.911, 1.935, 1.998, 2)</td>
<td>14.81</td>
<td>-6.34</td>
<td>3.07</td>
</tr>
<tr>
<td>(0.48, 105, 146, 175, 190, 213, 248), (0.832, 1.488, 1.622, 1.721, 1.737, 1.814, 1.899, 1.972)</td>
<td>9.30</td>
<td>6.25</td>
<td>1.35</td>
</tr>
<tr>
<td>(0.57, 91, 110, 170, 194, 246, 299), (0.179, 1.741, 1.947, 1.947, 2.2, 2.2, 2, 2, 2)</td>
<td>-3.41</td>
<td>5.99</td>
<td>0.68</td>
</tr>
<tr>
<td>(0.19, 220, 256, 290, 300, 300, 300), (0.011, 1.622, 1.656, 1.729, 1.912, 1.985, 1.901, 1.992)</td>
<td>2.73</td>
<td>4.33</td>
<td>0.86</td>
</tr>
<tr>
<td>(0.32, 83, 158, 247, 282, 286, 295), (0.433, 1.983, 1.985, 2.2, 2.2, 2, 2, 2)</td>
<td>-7.83</td>
<td>16.51</td>
<td>0.77</td>
</tr>
<tr>
<td>(0.45, 103, 117, 181, 288, 299, 239), (0.477, 1.226, 1.727, 1.847, 1.961, 1.986, 1.996, 1.997)</td>
<td>6.11</td>
<td>-5.89</td>
<td>-0.10</td>
</tr>
<tr>
<td>(0.32, 36, 107, 209, 246, 282, 271), (0.389, 0.673, 1.763, 1.925, 1.952, 1.979, 1.989, 1.999)</td>
<td>-2.88</td>
<td>11.71</td>
<td>-0.12</td>
</tr>
<tr>
<td>(0.27, 162, 187, 295, 298, 300, 300), (0.389, 1.584, 1.982, 1.984, 1.991, 1.996, 1.996, 1.999)</td>
<td>-3.70</td>
<td>5.97</td>
<td>-0.40</td>
</tr>
</tbody>
</table>

Figure 8. Details on the simulated policies

The table lists the policy, identified by a vector of the brackets and the price therein, the difference in water consumption (in %), the difference in revenues when no difference in consumption is taken into account (in %), and the expected revenues taking into account the changes in water consumption (in %).

Figure 8 shows a screenshot of the output table. There are filters for sorting the results on the different performance indicators. Currently, the table is sorted by the final revenues. The discrepancy between the different revenue values shows that the changing a pricing scheme while ignoring the elasticity in water consumption has potential threats. The ninth policy shows that in the pricing scheme one would expect an increase in revenues if one ignores consumer behavior, however, in practice one would observe a decrease in revenues. Note that the fourth policy achieves both a reduction in water consumption while at the same time the revenues are increased.

5. Technical validity of the adaptive pricing DSS

Taking into account the primary target of the project – the reduction of water consumption and balancing the demand with available resources - the methodology
is focused on the assessment of the water savings using the economic instrument (water tariffs). The general approach can be described as:

\[ dQ = f(dP), \]  

(1)

where \( dQ \) represents changes in the quantity of a raw good or services (water supply in this case), and \( dP \) represents the change of the price for the good/service.

Of course, there are many other factors influencing changes in the quantity of consumed water, but this module is focused on economic instruments. In more details, but still focusing only on the economic aspects, the following formula describes the responsiveness of the quantity of a raw good or service demanded to changes in its price:

\[ dQ = e_p \cdot \frac{dP}{P} \cdot Q, \]  

(2)

where \( e_p \) is the coefficient of price elasticity of demand, \( P \) the initial price, and \( Q \) the quantity demanded before the changes of the price, in the calculation period.

In case of water demand, the term “price” is more complicated, especially, for mixed tariffs that consist of a flat rate and a volumetric charge. Therefore, the real price of water is derived using the formula below:

\[ P = \frac{E_x}{Q}, \]  

(3)

where \( E_x \) is the total expenditure per client in the calculation period, \( E_x = FL_f + VCh \) with \( FL_f \) the flat rate in the calculation period, and \( VCh \) the volumetric charge (quantity or quantities - multiplied by price or prices). The desk review gives us the estimation of the \( e_p \) in local conditions, the other data was derived from water operators.

Such methodology creates the possibility of predicting water saving by introducing changes in water tariffs. Such effects are related to an increase of water tariffs, however, the constructed model allows to review and check the affordability aspect.

6. Conclusion

The dynamic pricing tools allow simulating different pricing schemes under different assumptions. It shows that the dynamic pricing problem is a hard problem in which a lot of care needs to be taken. The policies that are simulated need to be judged together with the information on the influences. One need to find a balance between water reduction, increase in revenues, and feasibility of the policy in terms of influences and fairness. The dynamic pricing tool assists in finding this balance and has the potential to be integrated into important pricing policy decisions at a
strategic level. It is important to remain in contact with stakeholders to show the benefit of adaptive pricing so as to implement it in practice and improve the validation.

REFERENCES


THE RECOMMENDATION ALGORITHM FOR AN ONLINE ART GALLERY

WALDEMAR KARWOWSKI *), JOANNA SOSNOWSKA, MARIAN RUSEK *)

*) Department of Informatics, Warsaw University of Life Sciences (SGGW)

The paper discusses the need for recommendations and the basic recommendation systems and algorithms. In the second part the design and implementation of the recommender system for online art gallery (photos, drawings, and paintings) is presented. The designed customized recommendation algorithm is based on collaborative filtering technique using the similarity between objects, improved by information from user profile. At the end conclusions of performed algorithm are formulated.

Keywords: algorithms, recommender system, collaborative filtering

1. Introduction

Nowadays most of web pages are created not only by the programmers and administrators but also by the users. This direction of the development of the Internet started at the beginning of 21 century with the possibility of commenting web content by users, and is known as Web 2.0 described in the Tim O'Reilly article from the year 2005 [1]. With the ability to publish and edit the content by any user, everybody can be a co-author of a portal. Main manifestation of this idea is the popularity of blogs, social networking and wiki services. Of course users’ impact on the content of the websites is not restricted to enable users to self-manage the content and appearance of the pages. Users provide their personal data and can be identified during their activity. This means that Internet services have a lot of information about users, for example history of activity or interest of topics. In the
age of the Internet, the customer is faced with the problem of excess offers and information. Similar situation appears, if user is looking for interesting information on the social networking site. Currently many systems provide mechanisms for automatically displaying personalized content on the basis of the data contained in the user profile and the history of his activity on the portal. Examples of such mechanisms are recommendations to help user finding interesting new content, services with similar or related topics and people interested in such topics. The most common applications of personalization mechanisms are media-sharing services such as videos (YouTube) or images (DeviantArt), thematic portals with reviews of movies (FilmWeb) computer games and books and, above all, e-commerce systems and online stores, of which the largest and most famous is the amazon.com. Such systems are called recommendation or recommender systems. However, recommendation systems have not appeared together with Web 2.0, they are much older. The first recommendation systems concepts already appeared before year 1980, much earlier than the first Internet portal. A prototype system of recommendations was “electronic Librarian” Grundy [2], book proposals system for reader, based on the information provided by the reader and the preset “stereotypes” about readers taste. The first real running, but partially manual, recommendation system was Tapestry [3]. It allowed the user to query for items in an information domain and had the task of filtering the documents, for example messages in internal e-mail systems used by corporations. The purpose of its use was to release users from the unnecessary messages. Shortly after it appeared fully automated filtering systems GroupLens [4]. It was locating relevant opinions automatically and aggregating them to provide recommendations to identify Usenet articles which are likely to be interesting to a particular user. Generally a recommendation process is closely related to the filtering of information, in mentioned examples, recommendation was designed to help user choose proper documents by filtering. Today, recommendation more frequently may be a way to offer the product to the purchaser.

The aim of the work was an implementation of the simple recommendation system for an online art gallery MyArtGallery. MyArtGallery is typical Web 2.0 service and was created in ASP.NET MVC 5 technology as a part of first degree thesis. The main functionality of the MyArtGallery is the ability to publish users’ work in various fields of the wider art. Main gallery functionalities are among other commenting on the work of other users, add images to own collection of favorites, download images in the selected resolution on own PC and much more. The images are divided into genres (categories) and are described with keywords (tags). The application also offers the ability to search for images based on specified criteria. Analysis of the first version of the MyArtGallery showed that a recommendation system would be useful. The recommendation system was needed to facilitate the user to discover images that may be of interest to him.
The rest of this paper is organized as follows: in Sect. 2 the concept of a recommender system, together with most important recommendation techniques and algorithms is presented. In Sect. 3 problems connected with recommendation for an online art gallery is discussed. In Sect. 4 an original recommendation algorithm for an online art gallery is precisely described. We finish with summary and brief remarks in Sect. 5.

2. Recommendation systems and algorithms

Assisting the user in making decisions is very important because of the widespread information overload. Information overload comes from the fact that modern man meets daily with much more information than he is able to process, i.e. understand and remember. This problem is much older than the World Wide Web; however thanks to the dynamic development of the Internet, it has become particularly disruptive. The consequence of the increasingly more widespread access to the Internet in all parts of the world in conjunction with the use of Web 2.0 philosophy is the fact that the amount of new information grows with the number of users. The book dedicated to the problem of information overload [5] describes yet another variation of this phenomenon-overload messages (the message) resulting from the popularity of new forms of communication such as social networking sites, post office email and mobile technologies. According to the Internet Live Stats (http://www.internetlivestats.com/one-second/), within each second it is published more than 7 thousand entries on the social networking site Twitter, on Instagram is published more than 700 new photographs, and is sent more than 2.5 million e-mails (October 2017). In this situation, the possibilities offered by traditional search engines are inadequate. The use of traditional search engines involves the necessity of independent browsing hundreds or even thousands of pages of results. Users need recommendations from trusted sources to make decisions; this means that information filtering systems are very important.

Definitions of recommendation systems are rather descriptive. According to [6]: “The goal of a Recommender System is to generate meaningful recommendations to a collection of users for items or products that might interest them”. Currently the most common are contacts between sellers and buyers and Recommendation Systems become one of the most powerful and popular tools in electronic commerce. In other words “Recommender Systems have evolved to fulfill the natural dual need of buyers and sellers by automating the generation of recommendations based on data analysis” [6]. It is possible because sellers and site owners have a large collection of data gathered about users that allows for deeper analysis of how a user interacts with topics, items etc. From the other side users need to personalize their online environment to overcome information overload. We can define Recommender systems as tools to help people make decisions in complex
information spaces [7]. According to [8]: “Recommender Systems are software tools and techniques providing suggestions for items to be of use to a user. The suggestions provided are aimed at supporting their users in various decision-making processes, such as what items to buy, what music to listen, or what news to read”. Most important functions of recommender systems are listed in [8]. From the service providers’ point of view there are: increase the number of items sold, sell more diverse items, increase the user satisfaction, increase user fidelity, better understand what the user wants. From the users’ point of view there are: find some good items, find all good items, annotation in context, recommend a sequence, recommend a bundle, just browsing, find credible recommender, improve the profile, express self, help others, influence others. In order to implement its core functionality, identifying the useful items for the user, a Recommendation System must predict that an item is worth recommending. In order to do this, the system must be able to predict the utility of some of them, or at least compare the utility of some items, and then decide what items to recommend based on this comparison [8].

Process of generating a recommendation depends on, among others, from the system destination, implemented functionality and availability of data on the user activities. Many recommendation algorithms use the history of reviews or other activities that could be construed as equivalent to evaluate items. According to [6] Recommender Systems can be broadly categorized as two types. In Collaborative Filtering systems a user is recommended items based on the past ratings of all users collectively. The second type is Content-based recommending where recommend items are similar in content to items the user has liked in the past, or matched to attributes of the user. Moreover [6] define many Hybrid approaches which combine both collaborative and content based approaches. In [9] recommenders systems are classified as collaborative filtering and knowledge-based approaches. Collaborative filtering is a real-time personalization technique that leverages similarities between people to make recommendations. In contrast, a knowledge-based recommender system exploits its knowledge base of the product domain to generate recommendations to a user, by reasoning about what products meet the users’ requirements. Much wider taxonomy was provided in [10] it distinguishes between four different classes of recommendation techniques based on knowledge source of recommendation approaches. Collaborative: the system generates recommendations using only information about rating profiles for different users. Content-based: the system generates recommendations from two sources: the features associated with products and the ratings that a user has given them. Demographic: a demographic recommender provides recommendations based on a demographic profile of the user. Recommended products can be produced for different demographic niches, by combining the ratings of users in those niches. Knowledge-based: a knowledge-based recommender suggests products based on inferences about a user’s needs and preferences. This knowledge will sometimes contain explicit functional
knowledge about how certain product features meet user needs. Additionally [8] distinguish community-based approach. This type of system recommends items based on the preferences of the users’ friends. Evidence suggests that people tend to rely more on recommendations from their friends than on recommendations from similar but anonymous individuals.

3. Recommendation problems for an online art gallery

The first version of MyArtGallery was implemented in ASP.NET MVC technology. The basic functionality of the MyArtGallery is the ability to publish users work in the field of visual art, for example: drawing, painting, photography and all other forms of art, which can be provided in the form of a digital image by photograph or scan. This ability is available only for registered users. The user account contains basic personal information such as, first name, last name, date of birth, gender, user name, e-mail address, avatar and information about interests. The publication process involves uploading image to the server from user computer or submission the image Web address, and entering in the information such as the title, short description of the image, the list of keywords and genre of art (category) from the selection list. The application has the ability to edit the information and image file and offers many additional functions such as commenting on the images of other users, create a collection of user’s favorite images, downloading graphic files in the selected resolution to user’s computer and reporting to administrator about images and comments which are illegal with the principles of the community. The functionality implemented in the first version also included the ability to search for images or user profiles according to the selected criteria. The search engine has two modes: simple and advanced.

![Figure 1. Advanced search. Source: own preparation](image)
Simple search means searching to find all the information about the images or profiles according to given word or phrase. The functionality of the advanced search is designed to allow users to discover art based on several different criteria. User fills the form (Fig.1) with the following criteria: the selected genre of art, whose description contains a given phrase, and is tagged by chosen keywords. Results can be sorted against date or popularity. However, both search methods have proven to be insufficient, because using the information in text form is associated with multiple disadvantages. First of all full text search is not always applicable because user can leave description field almost empty. Keywords (tags), and categories allow user to quickly search for images without the use of expensive computationally intensive searching text algorithms. However, this does not eliminate the underlying problem resulting from the application of a classic search engine, which is the need to accurately determine the search criteria that requires good orientation in the topic (e.g., frequent links between the keywords). In addition, image tags in the MyArtGallery are supplied by users; connecting image with keywords is very subjective. One user may provide significant keywords but the other rather random one. Another element to be taken into account is the problem of the different priorities of individual users, for one user more important is what the image shows and for another technique, in which item was made.

Mentioned above problems have resulted in the need for adding the recommendation system. The main goal was primarily to facilitate the user to discover interesting images and thus broaden his interest in the arts. The choice of algorithm for art gallery is a special challenge due to the nature of the published content, because art items evaluating is very subjective. For example, user might not like the illustration connected with favorite book drawn in a style that user does not like. Recommendations based on a single image should contain the art of similar themes and genre, but yet unknown for the currently logged user. Recommendations for a specific user should combine items similar to those that he already knows, and new for him but often liked by users with similar taste. It is also important to maintain a balance between current user interests and topics interesting him in the past.

Content-based solutions work well for many kinds of content. However, the use of a description of the image and keywords is very inaccurate for image recommendations. In addition, this method is very sensitive to errors such as mis-spelled category. Contrary common filtering recommendation algorithm based on the relationship between images makes the quality of the recommendations independent of the image description. However, this solution has a few restrictions particularly undesirable in the case of an art gallery like favoring the most popular images and popular types of art. Such recommendations are not good for users interested in niche genres of art. This problem is particularly visible for systems with a relatively small amount of data in the database.
The best choice was to create a hybrid solution, which combine the diversity of recommendations and their compatibility with the user's current interests. Recommendation system for the MyArtGallery is designed on the basis of the common filtering algorithm extended with the concepts of knowledge-based algorithm [11]. Common filtering enables us to generate a recommendation without relying on information about the images. Algorithm based on the knowledge compensates the basic restrictions associated with the common filtering. A solution of this type works well in e-commerce systems [12], and one of the objectives was to try the similar solution in the noncommercial social media system. It was decided to implement two types of recommendations: images similar to the currently displayed image and recommendations for the user, based on his favorite images. Both algorithms are influenced by the solution described in [13]. They are based on generated list of neighbors i.e. images often added to favorite lists by users interesting in in the past a certain image or group of images. List of images similar to the currently displayed image is created on the basis of the classic list of neighbors i.e. images with the highest similarity values to the currently displayed image. Recommendations based on user's interest are selected from lists of neighbors for each user images and from a list of his favorites.

4. Recommendation algorithm for an online art gallery

The process of generating recommendations for the user is divided into three basic steps. The first of these is to create a list of candidates. Then a list of “neighbors” for a sample image or a list of images is created with the use of common filtering based on the items. At the end the received recommendations are ordered on the basis of the values of similarity and additional information. The first step, initial filtering data, consists on preparing data for analysis and includes the creation of a list of candidates or images to be taken into account by the algorithm. Finally a list of all the images in the Gallery is limited by removing user paintings, his list of favorites and such, for which recommendations were earlier rejected by the user. The second step, creating a list of neighbors and a choice of n top recommendations is performed as the following. In both types of recommendations, the recommendations are generated using common filtering based on the links between images. The difference is how to create a list of neighbors and criteria for the selection of the n best results (it was set for MyArtGallery n = 20). Instead of the classic concept of the selection of the best neighbors based on the value of the similarity, recommendation algorithm for the MyArtGallery is using several criteria to organize potential recommendations. In addition to the value of the similarity, algorithm takes into account information about the categories and tags supplied by the user, so it can be referred to a hybrid solution that combines the elements of the com-
mon filtering and the concept of knowledge-based recommendations. This approach allows user to generate a satisfactory recommendation, regardless of the amount of own work and his favorites and number of links with existing user interests.

First type of recommendations is based on one sample image. When a user displays a page of the selected image, below the image a list entitled “similar images” is presented. Recommendations for a single image are not stored in the database, it is created dynamically. The first step of the algorithm is to create a user ratings matrix (see table 1).

<table>
<thead>
<tr>
<th>Sample Image</th>
<th>Image1</th>
<th>Image2</th>
<th>Image3</th>
<th>Image4</th>
</tr>
</thead>
<tbody>
<tr>
<td>User 1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>User 2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>User 3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>User 4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

The rows of the matrix correspond to users whose list of favorites contains the sample image, columns – images from the list of candidates added to favorites by at least one of user. If user i added an image j to own favorites, rating at the intersection of the row and column has the value 1, otherwise it has the value 0. The next and most important step of the algorithm is to generate a list of neighbors for the image. For each column of the matrix of ratings, similarity to the sample image is counted. As a measure of similarity is the Jaccard coefficient was chosen for vectors of likes of comparing images i.e. columns of the matrix. The primary argument for the choice of Jaccard measure of similarity was the fact that in the system using the favorites list “ratings” are binary values. Compared sets have the following form:

• column of sample image - consists of all ones, the number of components is equal to the number of users who have added the image to their favorites.
• column of the ratings matrix for j-th image.

The common part of the sets (columns) are components that both have a value of 1. To compare images based on interest history it is enough comparing the number of ones in the corresponding columns of the matrix. The pattern in the image of j-this image simplifies to equation 1:

\[
\text{similarity (j)} = \frac{\text{number of values 1 in column j}}{\text{number of sample image likes}}
\]  

(1)
The undoubted advantage of this measure of similarity is relatively small computational complexity - for each image it is perform one division, and addition at most as many “likes” has sample image. Because column of sample image likes always consists of all ones, the value in the denominator is the number of users who have added the image to their favorites. The identifiers of the images and the corresponding values of similarity are stored in a dictionary of potential recommendations and sorted in descending order and finally narrowed down to the 2n elements with the highest values of similarity. Thanks to that data analyzed in the further part of the algorithm is much smaller.

The next step is to create on the basis of the dictionary of potential recommendations new dictionary which assigns each identifier an array of numbers. These numbers are measures of the similarity of the two images: one image i contained in the dictionary of potential recommendations and the sample image.

For an identifier i array contains:

- the value of the similarities (Jaccard) for the image with identifier i from the dictionary of potential recommendations,
- the ratio of the number of common tags for images being compared to the number of all image tags with identifier i,
- a value of 0 or 1 represents the membership of comparable images to the same category of art (this value is also 1 if the image belongs to one of the subcategories of the category image sample).

This dictionary is first ordered in descending order by the common tags and limited to the n elements corresponding to the images with the highest content of common tags. Then the resulting dictionary is ordered according to the category. If images have the same content of common tags, as the first positions are images in the same category of art, what the sample image. For the image, which nobody has added to the favorite measure, dictionary contains the identifiers of all images from the list of candidates and tables consisting of only two values corresponding to the common tags and belonging to the category. List of recommendations appears in view is the list of images from the list of candidates, whose identifiers are included in the ultimate recommendations dictionary. The order of the images is the same as the order of the elements of the dictionary.

Recommendations based on the profile of the logged-on user are generated on the basis of a list of user’s favorite images, and his own images. They are generated each time the user navigates to the home page of the application. Before the start of the algorithm, existing recommendations for a particular user, in addition to the recommendations rejected by user, are removed from the database. The basis of the algorithm is a comprehensive dictionary of potential recommendations, initialized as an empty dictionary, similar to the dictionary for a single image. Then, for each image from the history of user activity is performed the following procedure. In
Step 1: A dictionary of similarities for the current image is created— the same way as for a single image. In step 2, the created dictionary is narrowed down to the elements with the highest values of similarity. In step 3, identifiers of images and assigned to them the values of similarity are added to the dictionary of potential recommendations. If the algorithm encounters an identifier that is already in the dictionary, the value of the similarity is summarized with the value assigned to the same key value. It means that the images similar for more than one image are more likely to find in the list of recommendations. Then in the same way second dictionary is created, with images potentially uninteresting for a particular user (anti-recommendations) based on images whose recommendations were rejected in the past by the user. It is limited to the n elements with the highest values of similarity. Items whose keys are included in this dictionary are removed from the dictionary of potential recommendations, and dictionary of potential recommendations is reduced to the 2n elements with the highest values of similarity.

Figure 2. User profile - preferences

The last stage of recommendation algorithm involves extracting the n best recommendation based on dictionary of potential recommendations and preferences from profile for particular user (Fig. 2). If user did not provide any additional information about his preferences, recommendations are generated based on the n items with the highest similarity values extracted from the dictionary of potential recommendations. If at least one of the four lists included in the profile of user interests is not empty, the new dictionary is created. Its keys are the image identifiers from the dictionary of potential recommendations and its values are arrays of numbers. The following conditions must be met to the identifier of the image found
in this dictionary: the image cannot belong to the category ignored by user and user
list of tags can contain a maximum of 10% of the tags that are marked by a given
user as uninteresting.

Array for an element with the identifier i consist of three elements:
• the similarity values for the image i,
• the number for the ratio of the number of image tags in the list of tags preferred
  by the user to the number of all tags for image i,
• a value of 0 or 1 which indicates a membership category of image i to the list
  of categories preferred by the user or their subcategories.

Extracting n the best recommendations from the dictionary is performed in the
same way as in the case of recommendations for a single image. If the list of simi-
larities is empty (for example, for a user who hasn't any own images, and nothing is
added to his favorites), the dictionary value is created based on the entire list of
candidates. Due to the lack of similarity values, an array for identifier i consist only
two numbers calculated based on user profile. The result is a list of images with the
highest content of the tags listed in the profile of user interests and preferred cate-
gories. The last stage of generating recommendations is to create recommendation
from the final version of the dictionary and save them in the database in the order
specified by the algorithm.

5. Conclusions and future work

The recommendation system has been tested by using 30 user accounts with
the sample data. Users’ interests have been chosen so way that some topics often
occur together (e.g. nature-photography, science fiction-fanart, traditional art-
portrait). Evaluating of the algorithm consisted of comparing how much the rec-
ommended images correspond to the interests of the user. The second aspect of
evaluating was to determine whether the topic of recommendations that go beyond
the current interests of the selected user is related to the common relationship be-
tween user preferences. For the first type of recommendation test were performed
for a selected set of examples. They have shown that the use of a recommendation
based on the relationship between users interested in a given image, even if the
image description is not very accurate, makes that the algorithm is able to find
images with similar themes. Using a tag description does not dramatically change
the results of a recommendation, but only increases precision. Tests for recommen-
dations based on the user’s profile have been tried for all user profiles created. It
was shown that profile-based recommendation functionality allows users expand
their interests and receive suggestions that match their current preferences. Howev-
er, for this type of recommendation it was necessary to clarify the user profile and
to reject the part of the proposed images. After such steps, the quality of the results was much better. It was concluded that the most important future improvement should be possibility of remember tags often encountered together. Remembered links can be used to generate hints of possible tags when a user fills out a preferences form.

REFERENCES

ASSESSMENT OF THE DEGREE OF FITTING THE TRANSPORT POTENTIAL OF THE TRANSPORTATION COMPANY TO A RANDOM DEMAND FOR TRANSPORT SERVICES

GUSTAW KONOPACKI

Military University of Technology, Department of Cybernetics, Institute of Information Systems

Will be considered transport company exploiting the uniform in the sense of destination, means of transport such as tankers, with a total transport potential equal to a loading units (e.g. tones). The company operates in the transport market, where demand for transport services is random. The problem of fitting the transport potential of a transport company to the demand for such services is formulated. As measure of this fit assumes that the probability of a firm's transport potential is not being exceeded by demand over the given time horizon. Practically useful formulas for estimating such probability for cases where demand for transport services is described by stationary and non-stationary normal stochastic processes. The results are illustrated by numerical examples.

Keywords: transport potential, demand, stochastic process

1. Introduction

Transport plays a significant role in every economy. This is due to the fact that it performs important service functions for other branches of the economy. There is a close relationship between the development of transport systems and economic development. Efficient transport determines economic growth, while its underdevelopment is becoming a barrier to economic development.
Transport is a sector of the economy that meets the needs of moving people and goods. Participants in the transport market are suppliers and buyers of transport services, who represent the supply of these services and demand for them. The proper management of a transport company requires knowledge of issues related to the behavior of transport market participants, i.e. [17]:
- transport company as a provider of transport services (supply side),
- charge carriers as the recipients of these services (demand side).

The volume of transport service supply is determined by [17, 18]:
- economic conditions - mainly prices and costs in the transport sector,
- technical conditions - mainly the number, type and quality of means of transport and the condition and quality of infrastructure in the transport sector,
- market conditions - supply-side market structures that may limit or facilitate the activities of carriers (e.g. alliances),
- administrative conditions - regulations limiting the possibility of carrying out transport (e.g. axle load in road transport, customs regulations).

Demand for transport services arises from transport needs, which are defined as the willingness or the necessity to move people or cargo from one place to another by means of transport [18]. In general, there are two types of demand:
- potential demand, which is understood the natural demand for the goods or services,
- actual, effective demand, which is understood as the ability to buy a good or service at a certain time for a certain number of buyers.

One of the basic problems of transport company management is to ensure its continuous presence on the market of transport services. This is achieved mainly by ensuring the appropriate transport potential of the means of transport available, in line with the anticipated demand for transport services.

Failure to match the transportation potential of the transport company to the demand for transport services that will shape the market for these services in the time horizon foreseen by company will, of course, lead to the following two situations:
- reduce the company's competitiveness in the transport services market, and even exit the market when transport capacity is lower than demand,
- incur additional costs for non-use of transport potential on the demand for transport services.

One of the measure of matching the transportation potential of a transport company to the demand for transport services may be the probability that in the forecast time horizon the transport potential will not be exceeded by demand - in the first situation or transport potential is exceeded by demand - in the second situation.

Usually, the first of the aforementioned situations will have greater implications for the transport company and will therefore be considered further,
taking into account the random nature of demand for transport services and will not take into account the operational characteristics of transport as was the case in [5].

2. Description of the problem

Consider transport company using the same or very similar means of transport (e.g. tankers) is used to meet the demand for homogeneous transport services (e.g. fuel transportation). It is assumed that the means of transport need not be the same i.e. they do not have to have the same construction solutions and the same capacity. The transport potential of the transport company will be the total capacity of all currently available means of transport. This potential will be determined by \( a \) and treated as the ability of the transport company to provide transport services at most in this size. Transport potential can be either constant or random. It is also assumed that the demand for transport services is random and can be interpreted as a stationary or non-stationary stochastic process. Thus, temporary demand values, which are the realization of random variables, may both be larger than and be lower than value of the transport potential of the transport company.

These cases have a character of temporal mismatch of transport potential to demand, which occur only in finite time intervals and the length of which depends on the probabilistic characteristics of the process of formation the size of demand for transport services in case the performance characteristics of the means of transport are not taken into account, as is the case in this article.

Next will be considered the second case, which seems to give rise to more serious consequences for the transport company and as previously noted, the lack of ability to meet demand can lead to a undermine of the transport company's position in the transport services market and even to its exit of the market. Therefore, it seems necessary previous discernment the transport company about its ability to function in the transport services market, if it has transport potential of \( a \). This knowledge can be useful for making decisions such as buying additional means of transport or changing the profile of providing transport services.

3. Demand for transport services of unchanging trend

3.1. Constant value of transport potential

It is presumed that the further demand for transport services can be described by means of continuous stochastic process \( X(t) \) class CC. It is assumed also that process \( X(t) \) is stationary, ergodic and differentiable in the mean-square sense [1−3, 6, 9, 13, 14, 16, 19]. Let \( m_x \) be the expected value of this process and \( K_x(\tau)=\sigma^2_r(\tau) \) its correlation function, where \( r(\tau) \) defines a normalized correlation function. The
problem to be considered is to determine the probability of exceedance by the demand for transport services (process $X(t)$) of the transport potential of transport company fixed at a constant level $a$. An exemplary implementation of demand for transport services as a function of time (exemplary implementation of the stochastic process $X(t)$) for the constant value of the transport potential is shown in Figure 1.

[Figure 1. Exemplary implementation of the stochastic process $X(t)$ describing the demand for transport services in case of constant value of transport potential]

This problem determining of the probability distribution of residence time of a stochastic process $X(t)$ over a defined threshold value $a$ is computationally difficult. Fortunately, however, in practice, often enough to know the expected value of the residence time of a stochastic process over a defined threshold value, which greatly facilitates the analytical solution.

The discussed problem was considered in detail in [6], and practically useful calculation formulas were obtained for the normal stochastic process in the form of the following estimates of the probability of not exceeding the transport potential of the transport company by the demand for transport services in the $T$-period:

- lower estimation

$$P_0(a,T) \geq P_0^{\min} = \Phi\left(\frac{a-m_x}{\sigma_x}\right) - n_0 \cdot T \cdot \exp\left(-\frac{(a-m_x)^2}{2 \cdot \sigma_x^2}\right)$$

(1)

- upper estimation

$$P_0(a,T) \leq P_0^{\max} = \Phi\left(\frac{a-m_x}{\sigma_x}\right) \cdot \exp\left(-n_0 \cdot T \cdot \exp\left(-\frac{(a-m_x)^2}{2 \cdot \sigma_x^2}\right)\right)$$

(2)
where \( n_0 \) is the expected number of heights by process \( X(t) \) of its expected value per unit of time, which can be taken as equal [12]:

\[
n_0 = \frac{1}{2\sqrt{\pi}}
\]

Estimating (1) and (2) can use if \( T \) satisfies the following inequality:

\[
T \leq \frac{\phi \left( \frac{a-m_x}{\sigma_x} \right)}{n_0} \cdot \exp \left( \frac{(a-m_x)^2}{2 \cdot \sigma_x^2} \right)
\]

Example 1
The transport company has the potential to transport equal to \( a = 6000 \) units (e.g. tons), while the demand for transport services market is described by a normal stationary stochastic process \( X(t) \) of the expected value \( m_x = 4000 \) units loading and correlation function \( K_x(\tau) = 640000 \exp(-2.56 \tau) \). Calculate the value of the lower \( (P_{\text{min}}) \) and upper \( (P_{\text{max}}) \) estimates of probability \( P(a, T) \) not exceedances of the value of the transport potential of the company (the ability to meet the demand for transport services in its entirety) by demand in the time horizon \( T = 1, 2, ..., 15 \) (e.g. months).

**Figure 2.** Graphical presentation of the lower \( (P_{\text{min}}) \) and upper \( (P_{\text{max}}) \) estimates for probability \( P(a, T) \) on the basis of the data of Example 1.
3.2. Random value of transport potential

The management of real businesses shows that the assumption of the constant transport potential of transport company is an optimistic assumption. Usually, this potential is shaped by the random factors characteristic of the process of exploitation of means of transport. It is therefore reasonable to consider the question of the degree of adaptation of the randomly changing transport potential to the random demand for transport services. Let the demand for transport services be described by the stochastic process $X(t)$ as in 3.1, but let the firm’s transport potential be a random variable with the normal distribution $N(m_a, \sigma_a)$. Let the transport potential and demand for transport services not will be correlated with each other.

Exemplary implementation of demand for transport services as a function of time (exemplary implementation of the stochastic process $X(t)$) for the random transport potential is shown in Figure 3.

![Figure 3](image)

**Figure 3.** Exemplary implementation of the stochastic process $X(t)$ describing the demand for transport services in case the random transport potential.

In this case, the expected number of exceedance by the process $X(t)$ of the random value of transport potential is expressed in terms of:

$$n_a = n_o \exp \left( -\frac{(m_a - m_x)^2}{2(\sigma_x^2 + \sigma_a^2)} \right)$$

Using a similar approach as in point 3.1, a lower and upper estimate of the probability of not exceeding the value of the transport enterprise’s transport potential value over a time interval of length $T$ is obtained.

$$n_a = n_o \exp \left( -\frac{r_0^2 \sigma_x^4}{\sigma_x^2 + \sigma_a^2} \right) \exp \left( -\frac{(m_a - m_x)^2}{2(\sigma_x^2 + \sigma_a^2)} \right)$$
These are:

- lower estimation

\[ P_0(a,T) \geq P_0^{min} = \Phi \left( \frac{m_a - m_x}{\sqrt{\sigma_x^2 + \sigma_a^2}} \right) - n_0 \cdot T \cdot \exp \left( - \frac{(m_a - m_x)^2}{2 \cdot (\sigma_x^2 + \sigma_a^2)} \right) \]  

(6)

- upper estimation

\[ P_0(a,T) \leq P_0^{max} = \Phi \left( \frac{m_a - m_x}{\sqrt{\sigma_x^2 + \sigma_a^2}} \right) \exp \left( - n_0 \cdot T \cdot \exp \left( - \frac{(m_a - m_x)^2}{2 \cdot (\sigma_x^2 + \sigma_a^2)} \right) \right) \]  

(7)

Estimating (6) and (7) can be used if \( T \) satisfies the following inequality:

\[ T \leq \frac{\Phi \left( \frac{m_a - m_x}{\sqrt{\sigma_x^2 + \sigma_a^2}} \right) \exp \left( - \frac{(m_a - m_x)^2}{2 \cdot (\sigma_x^2 + \sigma_a^2)} \right)}{n_0} \]  

(8)

Example 2

The transport company has a transport potential that is a random variable with normal distribution \( N(6000, 400) \) units loading, while the demand for transport services is described by a normal stationary stochastic process \( X(t) \) of the expected value \( m_x = 4000 \) units loading and correlation function \( K_x(\tau) = 90000 \exp(-2.56 \tau) \). Calculate the value of the lower (\( P_0^{min} \)) and upper (\( P_0^{max} \)) estimates of probability \( P(a,T) \) not exceedances of the value of the transport potential of the company (the ability to meet the demand for transport services in its entirety) by demand in the time horizon \( T = 1, 2, ..., 15 \) (months).

Figure 4. Graphical presentation of the lower (\( P_0^{min} \)) and upper (\( P_0^{max} \)) estimates for probability \( P(a,T) \) on the basis of the data of Example 2
4. Demand for transport services of changing trend

The real market for transport services is usually characterized by demand, which is described by non-stationary continuous stochastic processes. This implies, in practice, that demand is shaped not only by random factors, but also by others that take the form of a trend. Further, formulas will be proposed to assess the fit of the transport potential of the company to the demand that is trended, which has a trend. These formulas - as in point 3 - will be express the probability of not exceeded the transport potential by the process describing the demand for transport services.

Estimation of the probability of not exceeded by the non-stationary stochastic process of the set threshold value $a$ can be calculated from the following general formula:

$$ P(a, t) \geq F_a(t_0) - \int_0^t n_a(\tau) d\tau $$

where: $F_a(t_0)$ - cumulative distribution function the coordinates of stochastic process at the initial moment $t_0$, $n_a(\tau)$ - expected positive number exceedances of the threshold value $a$ by the stochastic process in a unit of time.

In the general case, it is the great difficulty is an analytical calculation of the expected number of exceedances the threshold value by the stochastic process in a unit time. In [12] there are given formulas for determining this value for a normal non-stationary stochastic process.

Let us consider the case of demand, which can be described by the non-stationary stochastic process $Z(t)$ of the form:

$$ Z(t) = X(t) + \varphi(t) $$

where: $X(t)$ - normal stationary stochastic process with zero expected value and correlation function of form $K_x(\tau) = \sigma_x^2 r(\tau)$, $r(\tau)$ - normalized correlation function, $\varphi(t)$ - some non-random monotonic function of time.

An exemplary realization of demand for transport services as a function of time (exemplary realization of stochastic process $Z(t)$ with liner function of trend) is shown in Figure 5.
From (10) comes that the non-stationarity of process $Z(t)$ is conditioned by the function $\phi(t)$, while its randomness - by the process $X(t)$. If $\phi(t)$ is a monotonically increasing function, then the lower estimate of the probability of not exceedances by process $Z(t)$ of the threshold $a$ can be estimated from the dependence:

$$P(a,t) = \Phi\left(\frac{a - \phi(t)}{\sigma}\right) - \int_{0}^{t} n_{a}(\tau) d\tau$$  \hspace{1cm} (11)$$

where

$$n_{a}(t) = \frac{1}{2\pi \sqrt{-r''(\theta)}} \cdot \exp\left\{-\frac{1}{2} \left(\frac{a - \phi(t)}{\sigma}\right)^2\right\} \cdot \exp\left(\frac{1}{2 \sigma^2 r''(\theta)}\right) + \sqrt{2\pi} \frac{\phi'(t)}{\sigma \sqrt{-r''(\theta)}} \Phi\left(\frac{\phi'(t)}{\sigma \sqrt{-r''(\theta)}}\right)$$  \hspace{1cm} (12)$$

In the simplest case of the linear trend $\phi(t)$, i.e. when

$$\phi(t) = b_{0} + b_{1}t$$  \hspace{1cm} (13)$$

the discussed probability can be estimated from the lower using the following formula:

Figure 5. Exemplary implementation of the stochastic process $Z(t)$ with linear function of trend describing the demand for transport services
Formula (37) can be used only if $P(a,t) \geq 0$, i.e. when $t \leq t^*$, which is the solution of the following equation:

$$
\Phi\left(\frac{a-b_0-b_1 t}{\sigma}\right) - \Phi\left(\frac{a-b_0}{\sigma}\right) - \Phi\left(\frac{a-b_0-b_1 t}{\sigma}\right)
$$

(14)

Formula (37) can be used only if $P(a,t) \geq 0$, i.e. when $t \leq t^*$, which is the solution of the following equation:

$$
\Phi\left(\frac{a-b_0-b_1 t}{\sigma}\right) = \Phi\left(\frac{a-b_0}{\sigma}\right) - \Phi\left(\frac{a-b_0-b_1 t}{\sigma}\right)
$$

(15)

Example 3
The transport company has the potential to transport equal to $a = 3600$ units loading (e.g. tons), while the demand for transport services market is described by the non-stationary stochastic process $Z(t)$ as (30) in which trend is described by linear function as (36) with parameters: $b_0 = 3500$ and $b_1 = 5$ units loading. Process $X(t)$ is a normal stationary stochastic process of the expected value $m_x = 0$ units loading and correlation function

$$
K_{xx}(\tau) = 400 \exp(-2.56 \tau^2).
$$

Calculate the value of the lower ($P_{\text{min}}$) estimate of probability $P(a,T)$ not exceedances of the value of the transport potential of the company (the ability to meet the demand for transport services in its entirety) by demand in the time horizon $T = 1, 2, ..., 15$ (e.g. months).

![Figure 6](image-url)  

*Figure 6. Graphical presentation of the lower ($P_{\text{min}}$) estimate for probability $P(a,T)$ on the basis of the data of Example 3*
Unfortunately, for more complex relationships describing the volatility during the trend of demand for transport services, it is not possible to obtain satisfactory analytical formulas that allow us to estimate the probability of not exceedances by the demand of the transport potential of transport company. In these cases, the general dependence (11) and the dependence to determine the expected number of positive exceedances of the threshold value \( \alpha \) by the stochastic process in a unit of time according to the formula \( \varphi(t) \) are used. Detailed proposals in this regard are included in [12].

5. Conclusions

Points 3 and 4 dealt the question of the degree of adaptation of its transport potential to the demand in the transport services market which is a practical issue for any transport company. The assessment of this degree of fit allows the company to determine its position on the transport services market within the time horizon the specified by her, assuming that the demand for transport services is also shaped by random factors. Adopting this assumption has led to the need to treat transport demand as a certain stochastic process.

The article proposes that matching the transport potential to transport demand will be measured the value of the probability that demand will not exceed a certain threshold value which corresponds to the transport potential of the transport company. Two main scenarios that may occur in the transport services market are analyzed: first - demand for transport services is constant, second - demand for services has a time-varying trend, described by a real function. With additional assumptions, the demand for transport services in the first case was described by the stationary normal stochastic process \( \mathcal{CC} \) class, while in the second case by the non-stationary normal stochastic process of the same class. Both cases were dealt with in points 3 and 4 respectively.

In point 3 two detailed cases about the transport potential are discussed. In the first it was assumed that the potential is invariable in time, while in the second that the potential is a random variable. The latter case is closer to reality, because transport potential is largely shaped by random factors such as:
- loss of means of transport due to failure (randomness), technical wear, economic inefficiency, etc.,
- purchase or hire additional means of transport.

The results obtained are illustrated by numerical examples that they are consistent with intuition.

Because of the extent of the problem and the degree of complexity, less attention has been paid to the other case when the demand for transport services is described by the non-stationary stochastic process (point 4). The degree of
The difficulty of this issue is so large that failure to obtain for a wide class of non-stationarity detailed analytical expressions to estimate the probability that transport potential of transport company not exceeded by of non-stationary demand. The exhaustive analysis of this problem should include the following cases of reasons for the non-stationarity of demand for transport services, which caused by the existence of a trend described by:

- another function of real time than linear,
- by the random function.

REFERENCES


The paper presents the changes that have taken place in the E-commerce market in Europe. The European B2C e-commerce sales have been steadily growing since 2011. Still, the growth rate has decreased the last few years, from 18.4% in 2011 to 13.3% in 2015 and 12% in 2016. The value of market in 2016 was €509.9 billion. In the category of B2C e-commerce sales, the UK was the clear leader in Europe. In 2015, the British in total spent €157.1bn online, which is more than France (€64.9bn), Germany (€59.7bn) and Russia (€20.5bn) combined. In addition, the share of the UK was more than twice as much as all the other countries outside the top 10 combined. It was found that E-commerce market developed faster than economy but was dependent from her.

Keywords: e-commerce, B2C, European market

1. Introduction

E-commerce is a part of e-business, which, on the other hand, is a part of e-economy. E-commerce is the process of purchasing, selling and exchange of products, services and information via computer networks [1, 2]. Thus, e-commerce activity includes sales over the Internet, on Internet auctions via Web pages other than stores and via electronic mail [3]. The development of e-commerce took place in the early 1990s as the Internet emerged. The potential of the new distribution channel was first used by mail order companies [4, 5]. At first, the network was used only to publish company and product information.
tion of the appropriate legislation allowed for commercial activity in the network. The first Web store was Amazon.com bookstore, launched in 1995 [6, 7]. The increased pace of development of e-commerce was associated with shifting of financial and human resources to this field of activity. Enterprises, which were skeptical at first, were forced by the competition to invest in e-commerce [8]. After year 2005, e-business developed very quickly. This was aided by development of mobile applications and improvement of the electronic distribution and payment systems [9].

E-commerce encompasses transactions concluded via the network, based on IP protocol and via other computer networks, while final delivery of the product may take place outside the network [10]. Other definitions state that e-commerce includes EDI, cable and satellite TV, electronic cards, telephone communication, Intranet and Extranet networks [11]. As for entities participating in e-commerce, the following systems have been identified: business-to-business (B2B), business-to-consumer (B2C) business-to-public (B2P), consumer-to-consumer (C2C), consumer-to-business (C2B) [12].

In the electronic world the customer and the merchant do not meet face-to-face, and the clients are more discerning with increased options and solutions available to them online. As customers embrace e-commerce their expectations about service, support, and how they make purchases are changing. In the B2C commerce exchange, a recent process of disintermediation (using the web platform) has created entirely new “direct channel” opportunities [13].

Electronic commerce has redefined the distribution of all goods which are (or can be made to be) digital, such as software, music, documents, images, movies, and other information. Hard goods must still be delivered by traditional transportation methods, but with the introduction of the direct channel of communication between consumers and manufacturers, the innovative manufacturer can mass-customize the production process by reacting to the demand differences of individual buyers [14].

2. Research methodology

The main objective of the study was to assess the functioning and changes taking place on the market of e-commerce in Europe. The detailed objectives included: presentation of the meaning of e-commerce in economy, of the stages of changes on the e-commerce market, specification of changes on the e-commerce market, depicting of correlations between e-commerce and economic condition of the countries. Research encompassed the period of 2011-2016. The source material includes foreign and domestic literature on the subject, numerical data provided by Eurostat, Ecommerce Foundation and Worldbank, reports and studies on
e-commerce. The work was developed using the descriptive, tabular and graphic tools, as well as the correlation coefficient. In order to determine the relationship between the economic condition in the country and the e-commerce market, the correlation coefficient was applied. The indicators describing the economic condition of the country included value of GDP, GDP per capita, value of import and export and value of consumption expenditure of households. On the other hand, the parameters describing the condition of the e-commerce market included the market value.

3. The results

The full potential of the European e-commerce market has not yet been reached. In 2015 about 57% of European Internet users shop online, but only 16% of SMEs sold online. Also, only 16% of consumers shop online from another EU country. The number of jobs created directly and indirectly by the B2C e-commerce sector is estimated at nearly 2.5 million in Europe, a figure that will grow with the on-going increase and penetration of the Internet in the European society, and the projected growth of (B2C) e-commerce.

The average Internet penetration in Europe increased to 75.3% in 2015. For the countries in the European Union, this rate is slightly higher, as 81.5% of their residents was connected to the Internet last year. In the top 10 of European countries in terms of Internet penetration, it is interesting to see that all Scandinavian countries are represented (Tab. 1).

### Table 1. Top 10 countries in terms of Internet penetration in 2015

<table>
<thead>
<tr>
<th>Countries</th>
<th>Internet access* (%)</th>
<th>Online population (mln)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>75.0</td>
<td>515.8</td>
</tr>
<tr>
<td>EU28</td>
<td>81.5</td>
<td>350.2</td>
</tr>
<tr>
<td>Top 10</td>
<td>94.7</td>
<td>153.9</td>
</tr>
<tr>
<td>Iceland</td>
<td>98.0</td>
<td>0.3</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>98.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Netherlands</td>
<td>98.0</td>
<td>13.6</td>
</tr>
<tr>
<td>Denmark</td>
<td>97.0</td>
<td>4.6</td>
</tr>
<tr>
<td>Norway</td>
<td>97.0</td>
<td>4.1</td>
</tr>
<tr>
<td>Finland</td>
<td>93.0</td>
<td>4.3</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>93.0</td>
<td>49.8</td>
</tr>
<tr>
<td>Switzerland</td>
<td>92.0</td>
<td>6.4</td>
</tr>
<tr>
<td>Sweden</td>
<td>92.0</td>
<td>7.4</td>
</tr>
<tr>
<td>Germany</td>
<td>89.0</td>
<td>62.9</td>
</tr>
</tbody>
</table>

*Sources: Eurostat, Ecommerce Foundation and Worldbank, 2016

*share of total population excluding people aged 0-14
While this top 10 mainly consists of countries from Western, Central and Northern Europe, the top 5 of the lowest Internet penetration solely comprises countries from the Eastern and Southern European regions (Tab. 2).

Table 2. Top 5 lowest Internet penetration countries in 2015

<table>
<thead>
<tr>
<th>Countries</th>
<th>Internet access* (%)</th>
<th>Online population (mln)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 5</td>
<td>60.4</td>
<td>103.0</td>
</tr>
<tr>
<td>Turkey</td>
<td>54.0</td>
<td>31.3</td>
</tr>
<tr>
<td>Ukraine</td>
<td>58.0</td>
<td>21.9</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>60.0</td>
<td>3.7</td>
</tr>
<tr>
<td>Romania</td>
<td>62.0</td>
<td>10.5</td>
</tr>
<tr>
<td>Italy</td>
<td>68.0</td>
<td>35.7</td>
</tr>
</tbody>
</table>

Sources: Eurostat, Ecommerce Foundation and Worldbank, 2016
*share of total population excluding people aged 0-14

The real growth rate of the European Union’s Gross Domestic Product (GDP) was 2.6% in 2015 (Tab. 3). This represented a significant improvement compared to the preceding year, as the GDP was 1.8% in 2014. In terms of the share of e-commerce in GDP, the UK is clearly above the rest, with a share of 6.1%. It is interesting to see that the Northern European countries score quite well in terms of eGDP. Denmark (4.4%), Finland (3.5%), Norway (2.3%) and Sweden (2.2%) are all in the top 10.

Table 3. Top 10 countries in terms of eGDP in 2015

<table>
<thead>
<tr>
<th>Countries</th>
<th>GDP at market prices (billions euro)</th>
<th>GDP per capita at market prices (euro)</th>
<th>Share of e-commerce in GDP (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Europe</td>
<td>17,591</td>
<td>26,948</td>
<td>2.6</td>
</tr>
<tr>
<td>EU28</td>
<td>14,633</td>
<td>26,911</td>
<td>2.8</td>
</tr>
<tr>
<td>Top 10</td>
<td>7,414</td>
<td>41,070</td>
<td>3.1</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>2,568</td>
<td>39,400</td>
<td>6.1</td>
</tr>
<tr>
<td>Denmark</td>
<td>266</td>
<td>46,800</td>
<td>4.4</td>
</tr>
<tr>
<td>Finland</td>
<td>207</td>
<td>37,800</td>
<td>3.5</td>
</tr>
<tr>
<td>France</td>
<td>2,184</td>
<td>32,900</td>
<td>3.0</td>
</tr>
<tr>
<td>Ireland</td>
<td>215</td>
<td>46,200</td>
<td>2.8</td>
</tr>
<tr>
<td>Czechia</td>
<td>164</td>
<td>15,600</td>
<td>2.6</td>
</tr>
<tr>
<td>Netherlands</td>
<td>678</td>
<td>40,000</td>
<td>2.4</td>
</tr>
<tr>
<td>Norway</td>
<td>351</td>
<td>67,600</td>
<td>2.3</td>
</tr>
<tr>
<td>Austria</td>
<td>337</td>
<td>39,100</td>
<td>2.2</td>
</tr>
<tr>
<td>Sweden</td>
<td>444</td>
<td>45,300</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Sources: Eurostat, Ecommerce Foundation and Worldbank, 2016
In 2016, retail e-commerce sales worldwide amounted to 1.86 trillion US dollars and e-retail revenues are projected to grow to 4.48 trillion US dollars in 2021.

The European B2C e-commerce sales have been growing steadily since 2011. Still, the growth rate has decreased the last few years, from 18.4% in 2011 to 13.3% in 2015. This trend was expected to continue in 2016 (growth 12%), resulting in a European B2C ecommerce turnover of €509.9bn (Fig. 1).

In the category of B2C e-commerce sales, the UK was the clear leader in Europe. In 2015, the British in total spent €157.1bn online, which is more than France (€64.9bn), Germany (€59.7bn) and Russia (€20.5bn) combined (Fig. 2). The UK market had a share of more than one third of the entire European e-commerce market. In addition, the share of the UK was more than twice as much as all the other countries outside the top 10 combined. The share of the top 10 countries in the total European B2C ecommerce (€383.0bn) markets is almost 85%. However, within the top 10 the difference between the United Kingdom, France and Germany and the rest was also significant. Together, the UK, France and Germany account for 61.9%, while the other seven countries combine for 22.3%. The differences between the seven smaller markets in the top 10 were almost negligible. The share of number four Russia (4.5%) was only 2.6% larger than that of number ten Switzerland (1.9%).

![Figure 1. Total online sales of goods and services in Europe in 2011-2016](image_url)

*Sources: Eurostat, Ecommerce Foundation and Worldbank, 2016*
Table 4 presents results of the Pearson correlation coefficient and p value. The limit value of significance level was assumed to be \( p = 0.05 \). Significant correlations were marked by grey background of the text. Correlation coefficients were calculated for years 2011-2016 and for 20 biggest national e-commerce market in Europe.

Table 4. Pearson correlation coefficients between the situation and changes in the IT market and selected parameters of the economy

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Pearson correlation coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>value of GDP</td>
</tr>
<tr>
<td>correlation</td>
<td>0.778</td>
</tr>
<tr>
<td>( p ) value</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Very strong positive correlation values were found to exist between the condition of economy and value of the e-commerce market. Only in the case of GDP per capita in relation to the value e-commerce market, wasn't correlation. The relations were weaker for changes in exports than imports. E-commerce market developed faster but was dependent from economy.
4. Summary

The European B2C e-commerce sales have been steadily growing since 2011. The market of e-commerce in Europe is among the fastest developing markets – the dynamics of development in years 2011 – 2016 exceeded ten percent of growth annually. The value of market in 2016 was €509.9 billion. In the category of B2C e-commerce sales, the UK was the clear leader in Europe. In 2015, the British in total spent €157.1 bn. online, which is more than France (€64.9 bn), Germany (€59.7 bn) and Russia (€20.5 bn) combined. In addition, the share of the UK was more than twice as much as all the other countries outside the top 10 combined.

The number of jobs created directly and indirectly by the B2C e-commerce sector is estimated at nearly 2.5 million in Europe. The key factor in the development of the e-commerce market in relation B2C is Internet access. The average Internet penetration in Europe increased to 75.3% in 2015. For the countries in the European Union, this rate is slightly higher, as 81.5% of their residents was connected to the Internet last year.

It was found very high level of correlation between the e-commerce market value and the economy indicators in years 2011-2016. No correlation was found when comparing the e-commerce market value and GDP per capita.

REFERENCES


ERP SYSTEMS AND WAREHOUSE MANAGEMENT BY WMS

TOMASZ WOŹNIAKOWSKI *) , PIOTR JAŁOWIECKI *) , KRZYSZTOF ZMARZŁOWSKI *) , MAGDALENA NOWAKOWSKA

*) Department of Informatics, Warsaw University of Life Sciences (SGGW)

This article presents differences between Enterprise Resource Planning Systems and Warehouse Management Systems. Management of warehouse processes and hardware requires specialized functionality, not present in ERP, functionality that only WMS can provide. Using separate systems enforces their integration. Different approaches to integration and risks connected to implementation are discussed and summarized.

Keywords: WMS, ERP, logistics, warehouse

1. Introduction

Logistics is one of those sectors which couldn’t effectively function without information technology support. The aim of the paper is analysis of integration issues among WMS and ERP systems in terms of costs and technology. Growing importance of competition in logistics market and continuous pressure on raising enterprises operations efficiency results in searching for new solutions and technologies, and increases importance of modern IT systems. The problem appears with choosing the right kind of solution that provides required level of efficiency for a reasonable costs. The necessity of choosing proper solutions is apparent on every stage of informatics investment. Implementing of such systems enables to improve enterprise financial condition, improve efficiency of functioning simultaneously in various areas, enables acquire full control over enterprise, and promises perspective of undisturbed development. By undertaking
implementation of integrated information system the company gets unique opportunity to make significant organizational changes. These operations may result in raising efficiency not only everyday work organization, but magnitude of results depends on scale and range of changes [1].

2. Enterprise Resource Planning systems

Enterprise Resource Planning (ERP) system is a package used for planning enterprise resources, composed of integrated modules which manage all of the core business processes of the organization. Main objective of the system is to seamlessly incorporate business processes within and across the functional and technical boundaries in the organization with improved workflow, standardization of business practice and access to up-to-date real time information. The basis of the ERP system is database layer, DBMS common for all the modules, that manages the operational and business data throughout the whole enterprise. The core functionalities include logistics management, manufacturing, sales and distribution, financial management, controlling, human resource management, project management, inventory management, service and maintenance, and management reporting.

ERP systems are very flexible and tailorable, allowing easily adaptation to specific requirements of different organizations because business functionalities are designed as independent components, which can function separately. Privileges system allows to define user access for individual employees only to resources that are related to their job’s responsibilities.

Basic functional modules of ERP system including logistics management consist of:

- Sales and Distribution – part of large logistics area, supports customer relationship from offer through the purchase order up to invoice. This module is closely integrated with inventory management and production planning.
- Materials Management – this module is responsible for the coordination of planning, sourcing, purchasing, moving, storing and controlling materials
- Production Planning - manages the lifecycle of product manufacturing process according to customer requirements
- Quality Management – module manages planning and controlling product quality, collects quality data of raw materials, manufactured goods and finished products.
- Plant Maintenance module for managing inspections, determining current technical condition of equipment, preventive maintenance and keeping technical equipment in perfect condition [2].
3. Main areas of risk by implementing ERP system

Every company which plans to implement ERP system have to take into consideration specific risk level. The number of analysed risk areas should be restricted to areas that have main impact on results of the implementation.

First is risk of not launching ERP system at all. This area describes a situation when as a result of different external reasons on side of IT system provider implementation was cancelled or not completed with a fully operational and commissioned deployment. Essentials of this risk area are expenses paid by the company for implementation ERP system. This cost should include not only initial offer cost but also expenses on engaging company employees in the implementation, or purchasing of servers and other IT hardware.

Another risk area connected with implementing ERP system in company is budget overrun. It is most common arising implementation risk, regardless of the class of integrated IT system. It is very hard to avoid this course of events. In a situation when a project is highly advanced the management rarely decides not to allocate additional resource for full system implementation. That’s why IT systems providers often underestimate implementation costs in theirs offer or omit important expensive elements of the system, like future users’ training or project administration, counting on adjudicate additional funding after process of IT system implementation will start.

Next area of risk in implementing an ERP class system is situation when company doesn’t achieve planned benefits from implementing ERP system. During making decision of implementing an ERP system, main and decisive factor is range of benefits for the company from an implementation of ERP system, That why this risk area should be considered as the most important amongst different risk areas. Probability of reaching planned business goals depends on functional sophistication advance of the system, which manifests inability to meet business requirements.

The last area of risk associated with implementing ERP informational system is delay in implementation. This risk area resembles risk of no achieving planned benefits from implementation of ERP system. It refers to situation when system ERP was fully implemented, assumed business profits was achieved, but with a considerable delay e.g. 3-6 months. Factors which influence this risk area are quality of project management in implementing company.[3].

4. WMS System

Warehouse Management System (WMS) is an informational system dedicated to manage entire high–volume warehouse operation in a real-time mode. These systems are made to optimize supervision on handling and storage of products.
Nowadays it is hard to imagine an efficient realization of logistics processes without IT tools that support them. The WMS program makes possible to quick receipt and pick up goods at the warehouse, defines an external supplier or specifies interior origin. It controls goods in terms of quality and quantity and automatically chooses storing location. The system specifies the shipping zone for orders marked for delivery, and analyses orders.

The consequences of using these system functions are increase in company turnover. Using such software makes possible to optimize tasks and realization of operations according to established priorities. As a result of implementing WMS system enterprise is supported in scopes of organization, supervision and control of processes related to movement and storage of goods and materials.

WMS have a number of specific features that accurately describe their behaviour. There will be

- optimization of storage space usage
- reduction of time for ordering and delivery of goods
- increasing of stock and assets turnover
- improving quality of services provided by manufacturers
- reduction amount of possible errors, due to advanced control and quick problems resolution between manufacturers and merchants
- high flexibility and mobility in exchanging data with the system
- easier data access
- full control of orders
- management of warehouse traffic
- assistance in preparation of dispatch documentation and automation of dispatch process
- automation of inventory process [4].

5. Differences between ERP Warehouse module versus Best-of-Breed WMS

Both large and medium size manufacturing and distribution companies are aware of the fact that functioning without WMS system is difficult in competitive market. As far as ERP systems become standard in enterprise management, WMS system also becoming more important.

Understanding the difference between WMS and ERP is very important for functioning of entire enterprise. Storage is last step during realization of customer’s order and determines the level of satisfaction from collaboration with product supplier.

WMS are system specialized to support realization of warehouse processes and warehouse management. They possess advanced functionalities, but also have high cost and long implementation time. Respondents who marked using of this
kind of solution were asked what was lacking in EPR warehouse modules implemented in theirs’ companies, that they decided to invest in separate WMS system. First they paid attention to greater flexibility of the WMS systems, possibility to dynamically manage pick locations and more efficient work with radio terminals. Surprisingly opinions of IT professionals and logistics have been different. Logistics most frequently pointed out more flexibility and dynamism of WMS systems, while the IT specialists opinions were more diverged and 40% had problem with answering this question and choose “hard to say” option.[5].

**Figure 1.** ERP warehouse module comparing to WMS
*Source: Systemy informatyczne w polskich magazynach, Raport 2014*
WMS during receiving goods uses advanced algorithms to optimize the place, time, and way to depositing goods, which is not supported by ERP. Logistics processes are accomplished by creating task queues that include various interleaving operations, minimizing the empty cycles of forklift trucks. ERP systems doesn’t work on queues, only on documents, which doesn’t contribute to increasing efficiency of work. [6][9].

| WMS | • if the company offers value-added services, such as kitting and assembly. It will need a specialized module for tracking this kind of work.  
• if the company clients are demanding greater visibility into inventory levels. Web portals that allow clients to view their inventory levels using a standard browser are still the province of best-of-breed solutions. |
| ERP | • If you need to improve reporting and accounting along with warehousing operations. ERP suites offer an enterprise-wide perspective on warehousing and supply chain operations. |

Source: own preparation based on https://www.softwareadvice.com/resources/wms-scm-erp-which-is-best-for-3pls/

6. WMS and ERP integration

Warehouse management does not exist in isolation. WMS often include integration with hardware: barcode scanners, printers, labelling systems with external shipping companies or integration with an internal ERP software.

Many warehouse operations, especially those with high-volume pick, pack and ship requirements, include sophisticated automatic material handling equipment which include conveyors, sorters, carousels, A-frame picking systems, pick-to-light systems. Most WMS solution offers extensive integrating with these devices.

Most manufactures have a need for integrated item information, customer orders, inventory, shipment and visibility of incoming product. The degree to which integration is necessary will vary depending on business needs. Integration with ERP implies that most software programs use a common database and most information is updated in real time. While there are many examples of ERP and WMS software systems that are truly integrated, most third-party WMS products are interfaced instead. Interfaced usually means the ERP system and WMS run on separate databases, and perhaps on separate servers. In this case, duplication of a
certain amount of business data is unavoidable, as well as additional implementation costs. Interfaces are usually accomplished with batch export and import programs that physically transfer data between systems, or a middleware product designed for this purpose. Even if these programs are included with the WMS or ERP package, during implementation may occur interface complexities and the need for custom interface development or interface issues that negatively impact the business.

Making decision about purchasing separate WMS and ERP systems, different integration scenarios were considered. Purchase the WMS package offered by ERP vendor that was originally acquired from a third-party vendor and is now offered as an optional bolt-on to the ERP package. This is a very common situation since, over the years, many ERP vendors have rushed to provide a WMS solution to fill this void. In order to make the WMS more compelling to potential customers, it is likely the new vendor spent plenty of time integrating or interfacing the system with its own ERP.

Purchase a best-of-breed WMS from a third-party vendor. Because it needs higher-end software capabilities, this may be the only valid option for many companies. However, the software will likely cost more to purchase and definitely
involves more custom interface expense. Also, using a middleware product to link the systems together will likely be a separate purchase [7].

Risk area connected with failed integration means that there will be significant amount of manual processing needed to keep data synchronized. So the easiness of integration, time and costs to integrate should be taken into consideration when choosing the WMS system

7. Using applications programming interfaces for integration

ERP and WMS manufactures are aware of the fact that their systems will be connected with other software. They offer different methods to extract and update data, including low-level programming languages in their software packages, providing Application Programming Interfaces (APIs), and documentation for database schemas. The most popular high level communication method is SOAP based web service.

Figure 3. UML diagram of basic middleware application
Web services use standardized XML messaging system to receive or send requests over the internet. Web services are platform independent, they are using standard protocols, can be implemented in any programming language. Definitions of specified service are published using WSDL (Web Services Description Language), an XML file which exposes service interface, interaction patterns and protocol mapping. WSDL can be easily interpreted by other applications, but it is hard to read by humans.

```xml
<operation name="add">
  <input name="addRequest" message="tns:addRequest" />
  <output name="addResponse" message="tns:addResponse" />
  <fault name="InsufficientPermissionFault" message="tns:InsufficientPermissionFault" />
  <fault name="InvalidSessionFault" message="tns:InvalidSessionFault" />
  <fault name="InvalidCredentialsFault" message="tns:InvalidCredentialsFault" />
  <fault name="ExceededConcurrentRequestLimitFault" message="tns:ExceededConcurrentRequestLimitFault" />
  <fault name="ExceededRequestLimitFault" message="tns:ExceededRequestLimitFault" />
  <fault name="ExceededUsageLimitFault" message="tns:ExceededUsageLimitFault" />
  <fault name="ExceededRecordCountFault" message="tns:ExceededRecordCountFault" />
  <fault name="ExceededRequestSizeFault" message="tns:ExceededRequestSizeFault" />
  <fault name="UnexpectedErrorFault" message="tns:UnexpectedErrorFault" />
</operation>
```

Extract from SuiteTalk WSDL, defining web service for Oracle NetSuite
Source: https://webservices.netsuite.com/wsdl/v2017_1_0/netsuite.wsdl

That’s why many programming platforms offers assistants, which can read WSDL file and create ready to use methods and classes. For instance Microsoft Visual studio enables adding “Web Reference” to project, automatically creates classes and methods using definitions from WSDL file.

Middleware applications use these automatically created classes to translate data received from one system to format required by other system, using predefined values for missing data where complexity of ERP system far exceeds range of information received from WMS system. Calling web service methods may fail, as a result of mismatched parameters, or unpredictable network problem, so all operation should be logged in local database.
8. Summary

WMS systems have considerably changed with regard to technology as well as amount of supported warehouse processes. For many years between the users of this systems classical approach prevailed – application should support typical processes for instance magazine’s registers. During the years users’ requirements grow. Companies wanted support in loading, picking, releasing, scanning, labelling and monitoring workers activities. Today we have to deal with situation when system have to be closely adjusted to clients requirements. Often it has to cooperate with industrial automation systems, because companies tend to fully automate their warehouses. Also grows importance of integration with data warehouses, carrier attendance applications, communications with subcontractors and stock management. More frequently WMS is becoming a part of the ERP system. Today integration of WMS solutions with advanced IT systems is relatively simple. ERP and WMS systems are becoming more flexible. Task that were impossible or very difficult few years ago, for instance implementing advanced ERP system in small company became common practice. At development of WMS system great influence has had mobile revolution, thank to which systems became easier to use – colourful application enables work with touch screen. Electronic information exchange in supply chain, mobile data update, integration with transport and warehouse automation have become standards.

The role of WMS also changes, it has to not only support warehouse management but also enable to create more intelligent forms of management. From this point of view WMS has to support more and more processes, it’s duties include scheduling tasks, allocating resources, managing transport automation and optimizing storage space. Today the role of the system is full automation of magazine’s work, including adaptation to external conditions. More important is mathematical support, for instance by using optimization algorithms. Of course, we have to remember that ERP never can replace WMS, just like ERP will not replace ERP system. There are two different solutions. ERP supports a number of business processes, including supply chain. WMS only manages the warehouse. They supplement each other, and their integration brings the enterprise most profit [8].

REFERENCES

[5] Systemy informatyczne w polskich magazynach, Raport 2014
[9] https://www.softwareadvice.com/resources/wms-scm-erp-which-is-best-for-3pls/