LONG TERM GROWTH OF SME FROM THE VIEW OF ICT COMPETENCIES AND WEB PRESENTATIONS

E + M

Klára Antlová, Luboš Popelínský, Jindřich Tandler

Introduction

The current business environment is characterized as highly turbulent, influenced by modern information and communication technologies, globalization, short innovation, production cycles and employee mobility [1], [15]. It is not easy for small and medium enterprises (SMEs) to compete in such an environment [11]. Therefore the organizations have to utilize their corporate resources to the greatest possible extent. Using information systems and particularly the web as dynamic tools can help companies to communicate with their customers and partners better and to improve their competitiveness.

Small and medium enterprises play an important role in all economies. They employ an increasing proportion of the total working population; bring new products and new services; they contribute to exports, national wealth, and competitiveness [19]. When compared with large enterprises, SMEs have a simpler organizational structure with fewer specialized tasks, poor human, financial and material resources. They also have a more limited use of information and communication technology (ICT) and the employees have fewer ICT skills. This low level of organizational readiness, insufficient skills and knowledge are the reasons for slow adoption of ICT. When the ICT is successfully implemented and used, the companies improve the following areas of businesses: reduced costs, fewer errors, reduced inventories, creation of new market opportunities, better communication and cooperation with customers and suppliers [1].

The successful implementation of ICT is one of the key factors of their long term growth. Some companies developed significantly more ICT competencies than others and these competencies help them to be more successful and competitive [16]. The authors would like to find if there is a correlation between ICT competencies (skills and knowledge) and a long term growth. First they would like to know which ICT competencies are the most important and have strong influence on the long term company growth. Secondly, the ICT competencies and probably the long term growth could be also associated with company web presentations. The authors would like to find out which attributes of the web page are the most important. This investigation will help SMEs to improve their competitiveness, better communication with customers and Internet marketing strategy. To summarize, the main goal of this contribution is to explore if there is a correspondence between ICT competencies of SME, their web presentation and long term growth. As a consequence, it can help an organization to improve employee education. To solve this task, the authors used various data mining methods, such as those described by Han & Kamber [8] and web mining methods [10] which were developed for solving classification and prediction tasks. They employed decision tree classifiers and association rules as well as class outlier detection methods and hierarchical cluster analysis.

A brief overview of relevant work can be found in the following section. Section 2 provides an introduction of the data mining method used in this work. Section 3 describes ICT competencies and data collection. Correspondence between long term growth and ICT competencies is explored in Section 4. The fifth part of the article explains the web mining methods and what can be extracted from web presentation. The last part summarizes the results.

1. Related Works

A number of authors [15], [17] see the **competency** as the knowledge, skills and attitudes used by employees in the performance of work. Such competencies require individuals or managers who posses capabilities in various areas. Hinckley and Perl [9] interviewed managers and established an understanding of competencies required by managers. They identified three areas of competencies: intra personal, interpersonal and organizational competency. As a part of their organizational competency is necessary

to cooperate and communicate with employees, partners and customers. ICT and also use of the Internet for communication and marketing play important role as a tool for these competencies.

Pepard and Ward [16] argue that ICT resources are combined (through structures, processes and roles) at the organization level to develop ICT competencies. They summarized six competencies which cover business strategy, information and IT strategy, exploitation, solution delivery and supply. These six domains contain together 26 competencies which can help an organization to use information system strategically and exploit its benefits. During the long term research the seventh competency was added by authors. This external competency expresses the ability to use software applications for EDI (Electronic Data Interchange). Long term research [1] documents that SMEs are under pressure from their partners to use EDI and this is very often the only reason why they innovate their ICT.

The ICT competencies refer to a firm's capacity to deploy their intangible resources, usually in combination with skills, to reach desired goals; this view is presented by Resource Based Theory [5]. This theory was developed to explain how organizations can achieve competitive advantage. Organizations' resources that are valuable and cannot be easily purchased require long learning process. Resource based theory explains why some companies evolved to become more successful than others.

Taylor [21] defines **web competencies** as the necessary skills and knowledge required for development of a web presentation. This means knowledge of many different areas: databases; programming; graphics; etc. The traditional skills of analysis/design and project management are also required, as they would be for any software project. As well as fundamental technical knowledge and skills, web developers also need to have skills in user interface design (also called human computer interfaces, or HCI); teamwork; communication (both written and oral); and they must be committed to life-long learning due to the rapidly changing nature of the field. The general business knowledge is also very important.

Cornish [6] and Buchner & Mulvenna [4] call the combination of company's Internet data and data mining techniques **marketing intelligence.** The data that is considered covers not only various types of server and web-meta information but also marketing data and knowledge.

Vescovi [23] describes collaboration between a university business school and several SMEs in Internet communication strategy development. In that process six typical problems were detected that may slow down the introduction of Internet communication. These problems were: unclear communication strategy, new communication paradigms, non integrating marketing communication, and company´ involvement in Internet challenge, organizational change, and people´ skills for Internet communication.

2. Data Mining Methods

For data analysis, data mining and web mining methods were used. These methods are based on various learning paradigms. We say with Tom Mitchell [14] that a machine learns with respect to a particular task T, performance metric P, and type of experience E, if the system reliably improves its performance P at task T, following experience E. Depending on how we specify T, P, and E, the learning task might also be called autonomous discovery, database updating, programming by example, etc. or data mining. For classification tasks, the goal is to classify unseen data into one of the predefined classes – in our case the task T is to classify a company as growing or non-growing – with the highest accuracy, i.e. the ratio of correctly classified examples, based on historical data, i.e. data seen before. Particular machine learning algorithms were described in [13] and [8].

Web mining [10] is usually split into two areas: web usage mining (when one mines from clickstream data) and web content mining when one learns from contents of web pages and also from pages that are referenced in a page, or reference this page. Here we focused on web content mining.

To analyze the data, Weka learning toolbox available at http://www.cs.waikato.ac.nz/ml/weka/ was used.

3. Data Collection

Data documented the long term company's economic results were obtained due to the cooperation between department of informatics (Business school in Liberec) and the coopera-

ting companies. This 15-year cooperation is based on the realization of the student's industrial internship. These companies host one-year student placements where students, as part of their Bachelor degree course, undertake long term work experience enabling them to integrate the practical and theoretical aspects of their course.

The thirty cooperating companies include the automotive, ICT, services and building industry. In these companies during the years 2008 and 2009 the ICT competencies were researched together with the company's managers, students (after one year internship) with the help of structured interviews. The questions contained

Tab. 1: ICT Competencies (Part 1)

Number	Area of ICT competencies
1.1	Business strategy has to identify uses of ICT
1.2	Incorporation of the potential of new technologies in long term business and ICT development
1.3	Establishing of appropriate criteria for decision making on investment in ICT
1.4	Defining of ICT management and the roles and responsibilities of information strategy
2.1	Prioritization (Ensure that the portfolio of investment in application and technology produce the maximum return from resources available)
2.2	ICT strategy alignment (Ensure that ICT development plans are integrated with organizational and functional strategic plans)
2.3	Business process design (Determine how ICT can deliver best practice in operational processes and organizational activities)
2.4	Business processes improvement (Identify the knowledge and information needed to deliver the strategic objectives through improved management processes),
2.5	System and process innovation (Carry out relevant R&D into how ICT can be used to create new ways of conducting business and new products or services)
3.1	Infrastructure development (Define and design information, application and technology architectures and organizations structures and processes to manage the resources)
3.2	Technology analysis resources (Understand technology trends and make appropriate recommendations for organizational acquisition of technology associated)
3.3	Sourcing strategies (Establish criteria and processes to evaluate supply options and contract with suppliers)
4.1	Benefits planning (Explicitly identify and plan to realize the benefits from ICT investment)
4.2	Benefits delivery (Monitor, measure and evaluate the benefits derived from ICT investment)
4.3	Managing change (Make the business and organizational changes required to maximize the benefits)
5.1	Application development (Develop/acquire and implement information, systems and technology solutions that satisfy business needs)
5.2	Service management (Define service arrangement and performance criteria to match business requirements including project management)

Tab. 1: ICT Competencies (Part 2)

Number	Area of ICT competencies
5.3	Information asset management (Establish and operate processes that ensure data information and knowledge management activities meet organizational needs and satisfy corporate policy)
5.4	Implementation management (Ensure that new processes and way of working are designed and implemented effectively in conjunction with new technology),
5.5	Apply technology (Deploy new/changed technology in the most cost effective mode to deliver application benefits)
5.6	Business continuity and security
6.1	Supplier relationship (Manage contracts and develop value added relationship with suppliers)
6.2	Technology standards (Develop and maintain appropriate standards, methods, controls and procedures for the use of ICT and associate resources)
6.3	Technology acquisition (Develop and apply procurement policies and procedures for the organizational acquisition of infrastructure components)
6.4	Asset and cost management (Ensure technology, information and application assets are effectively maintained and cost of acquisition and ownership are understood and managed)
6.5	ICT staff development (ICT staff development – recruit, train and deploy appropriate staff and ensure technical, business and personal skills meet the needs of the organization)
7.1	Customers relationship (Developing and maintain the relationship with customers and suppliers using EDI)

Source: [16]

the companies´ trends related to the all competencies mentioned in table 1.

All companies (in Tab. 2) are arranged according to their size (the number of employees: they have more then 9 and fewer than 200 employees), long term growth, their organizational development and ICT adoption. SME growth can be viewed through **five stages** of a company development [1], [11]. This model is based on the hypothesis that SME grows from small entities that are managed closely by the owner or the manager to larger ones. When the companies are growing the management style and using of ICT are changing. From our research we can also see that in the bigger and more successful companies the ICT competencies are presented and also more increasing. The researched companies were divided according above mentioned model. In Table 2 the existence of the competency within organization is indicated with "X", "-"means that competency does not exist.

The first stage of growth – **existence** is characterized by this business strategy: staying alive (the companies numbered 1–4). The company is working hard to find customers and to deliver orders. The organization is relatively simple, strategic decision is made in short term horizon and there is no strategic planning. There is minimal investment in ICT. As the business continues, the owners find they have to spend more time on administration. So they have to focus on profitability and improving administration processes. This leads owners towards the next stage.

The second stage – **survival** (the companies numbered 5–6) is characterized by establishing the company and growing the customer base. In this stage the information systems are simple: spreadsheets, mails, database of customers. The owner finds it more difficult to manage the growing number of staff. There is also a need for better management information for monitoring the market.

Tab. 2: Researched Companies

	IS competenices:																											
Org.	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	4.1	4.2	4.3	5.1	5.2	5.3	5.4	5.5	5.6	6.1	6.2	6.3	6.4	6.5	7.1	Grow
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	No
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	No
3	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	No
4	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	No
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	No
6	X	-	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-	-	-	-	-	-	X	No
7	X	X	-	-	X	X	-	-	-	X	-	-	X	-	-	X	-	-	-	-	-	-	-	-	-	-	-	No
8	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	No
9	X	X	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	X	-	-	-	-	-	-	No
10	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	No
11	X	-	-	-	-	-	-	X	-	-	-	-	X	-	-	-	-	-	-	-		-	-	-	-	-	-	No
12	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-		-	-	-	-	-	-	No
13	-	-	-	-	-	-	-	X	-	-	-	-	X	-	-	-	-	-	-	-		-	-	-	-	-	-	No
14	X	X	X	X	X	X	-	-	X	X	-	X	X	-	-	X	-	-	-	-	X	X	-	-	-	-	X	No
15	X	X	-	-	-	-	-	X	X	-	-	-	X	-	-	-	-	-	-	-		-	-	-	-	-	-	No
16	X	X	-	-	X	-	-	X	-	X	X	-	-	-	-	X	-	-	-	-	X	X	-	-	-	-	X	Yes
17	X	X	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		-	-	-	-	-	X	Yes
18	X	X	-	-	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-		X	X	-	-	X	X	Yes
19	X	X			X	X	X	X	-		X	-	X	-	X	X	-	-	X	-		X	X	X	-	X	X	Yes
20	X	X	X	-	X	X	X	-	X	X	-	X	X	-	-	X	-	-	-	-		X	X	-	-	X	X	Yes
21	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	X	X	-	X	X		X	X	X	X	X	X	Yes
22	X	X	X	X	X	X	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	X	X	X	-	X	X	Yes
23	X	X	X	X	X	X	X	X	X	X	X	X	X	-	X	X	X	X	X	X	X	X	X	X	X	X	X	Yes
24	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	X	X	X	X	X	X	X	X	X	-	X	X	Yes
25	X	X	-	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Yes
26	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Yes
27	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Yes
28	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Yes
29	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Yes
30	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Yes

Source: own

The third stage – **success** (the companies numbered 8–13) is where the company is well established, owners need to have a vision about their future, and to have bigger control of the company. Information systems are more important for managing of expanding customer base and increasing complexity of business processes.

The forth stage – **expansion** (the companies numbered 14–19) requires great skills of planning and leadership. The owner often recognizes the need for information system to improve communication and access to data within the company.

The last stage – **maturity** (the companies numbered 20–30) occurs when the company has grown to the size which has to be directed by the management team. Strategic planning is well established. The company has good relationship with their customers and suppliers, and ICT is strategically aligned with business strategy.

4. Relationship between Long Term Growth and ICT Competencies

4.1 Classification and Rule Extraction

The main task was to analyze how long-term growth is influenced by a particular competency (its presence or absence) or by a combination of competencies. At first it was necessary to learn whether there is any dependence between competencies and the long-term growth of a company. To answer that question the authors explored the data displayed in Table 2. Enterprises were characterized by 27 attributes (i.e. competencies as described above) where 'X' signs for presence and '-' for absence of the competency, and then classified into two classes - long-term growth (Grow=Yes) and no long-term growth or stagnation (Grow=No).

In this classification task the authors used decision tree learner J48 [8], [13]. The reason is that a decision tree is intelligible and also easy to analyze. As a quality criterion we used an overall accuracy, i.e. a percentage of correctly classified enterprises. For testing, because of a small number of examples, leave-one-out method [13] was used. The classifier was learned on 29 examples and then tested on the remaining one, and this was repeated 30-times, for all permutations. Then the average of these 30 runs has been computed.

The overall accuracy was very high and reached 83 %. It means that information about presence or absence of ICT competencies highly corresponds with company long-term growth. The competencies which are connected with information strategy have the key importance. Prioritization (2.1), ICT strategy alignment (2.2) and business process design (2.3) were the only competencies that appeared in the decision tree. The overall accuracy remained the same even if we removed 2.2 and 2.3. After removing also the competency 2.1 the long-term growth

depended only on technology analysis resources (3.2) and the accuracy slightly decreased to 80 %.

When analyzing the tree and looking for the most interesting paths in the tree, several decision rules were discovered that are displayed below.

It was observed that in the case that prioritization of ICT competency was missing (the competency 2.1="-") then the company was always non-growing. It holds for 13 companies. On the other hand, the presence of competency 2.1 appeared mostly in long-term growing enterprises but also in few not growing.

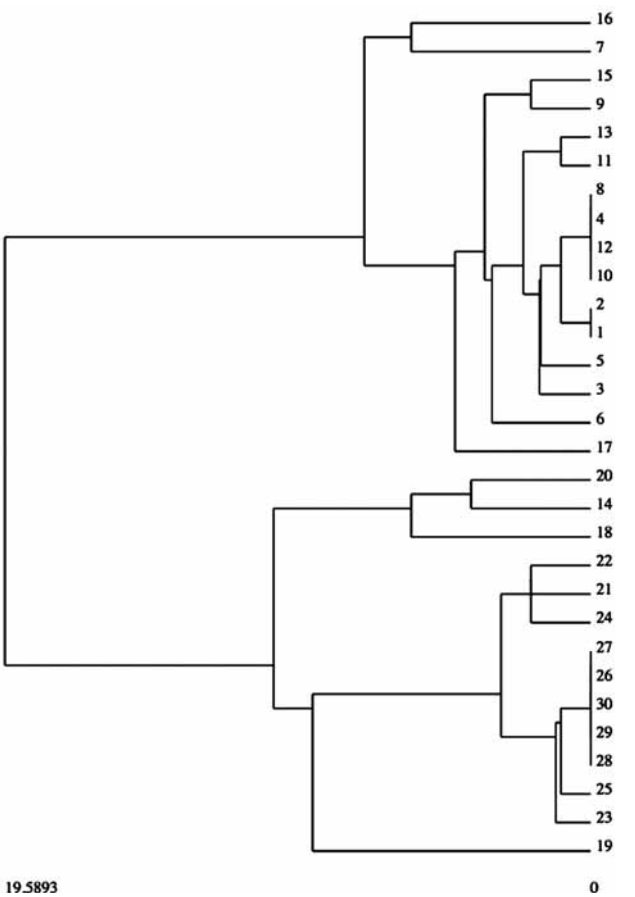
The authors also discovered that competencies 1.2 (ICT development) and 1.1 (business strategy) had the same effect as competency 2.1, i.e. missing competencies 1.2 and 1.1 arose in non-growing companies. On the other hand, the mutual presence of 1.2 and 1.1 does not imply such a strong dependence on long-term growth as 2.1.

It was also observed that the presence of competencies 2.3 (business process design), 3.2 (technology analysis) and 6.2 (technology standard) always appeared in a growing company and it holds for 13 SMEs. Similar dependence can be observed for competencies 4.3 (managing changes), 5.3 (information asset management), 5.4 (implementation management) and 6.3 (technology acquisition) but the number of growing enterprises is lower.

4.2 Introduction of a New Class

Although the classification accuracy was high, there were still almost 20 % of enterprises incorrectly classified. To discover what may be the

Fig. 1: Dendrogram



Source: own

reason we used clustering, namely hierarchical cluster analysis. We employed OC, a simple implementation of hierarchical clustering [3]. a distance between clusters was computed as a mean distance between their elements.

In the Fig. 1 the numbers on the right side identify a particular enterprise. Please bear in mind that the first fifteen enterprises (1–15) were classified as non long-term growing and the rest (16–30) as growing. a dendrogram is expected to cluster the most similar examples first. 5th hold perfectly e.g. for enterprises {1, 2} or {26–30}. On the other hand there are enterprises that are found similar and yet they belong to different classes – see cluster {7, 16} and {14, 20}. There is even the enterprise number 17 that belongs to a huge cluster that contains 13 non-growing enterprises. It means that company no. 17 is an outlier.

Clustering may suffer from the fact that all attribute (competencies) are taken as equally important. In fact we did not have any information that would allow us to set a weight (importance) of a particular competency. Thus to approve the observation above we used two outlier detection methods. Weka-CODB is a distance-based method for classified data [13]. Torgo [24] described another method that employed a hierarchical cluster analysis for ranking outlying values. With both methods we found the same set of outlying companies – 7, 14, 16 and 17.

Then we introduced a new class 'unknown', which contained enterprises 14, 16, 17 and also 7 and again employed the decision tree learner. For this three-class problem, accuracy increased up to 100 %. Now all non-growing companies are characterized either by an absence of prioritization or an absence of business process strategy although they display prioritization. Growing companies have both prioritization and business process strategy. If we compared this result for three-class problem to the above mentioned decision rules for long-term growth, we can see that competency 2.2 ICT strategy alignment is no more important for classification.

4.3 Association Rules

Association rules [8] are very useful for finding strong dependencies that are local, i.e. hold only for a subgroup of data. An association rule A => B says that if A holds than frequently B holds, where A (antecedent) and B (consequent) are conjunctions of items or events. Strength for an association rule is most often characterized by two measures, support, i.e. a fraction of data for which both A and B holds, and confidence, i.e. a relative frequency of finding B in data for which A holds.

For classified data, as in our case, a class association rule is a rule with only one item, class attribute, in the consequence. We looked for strong class association rules for the same classes as in the previous section – a growing/no growing company. We used Tertius [7] algorithm from Weka. The most interesting association rules are below.

For 74 % companies if the potential of new technologies is incorporated in long term business development then the company is successfully growing.

For 93 % companies if the potential of new technologies is incorporated in long term business development and if the relationship with customers is developed then the company is successfully growing. We could also observe a strong relationship between ICT infrastructure and developing a good relationship with suppliers.

For 93 % companies if the company does not have good technology analysis resources (3.2) and does not follow technology standards (6.2) then the company is not growing.

For 93 % companies if the company does not have good technology analysis resources (3.2) and does not have appropriate ICT staff (6.5) then the company is not growing. On the other hand the successful companies are looking for solutions that suit their customers and suppliers requirements. Insufficient ICT knowledge and ICT solutions delivery can be one of the reasons for the company stagnation. These problems we can find in 75 % of collected companies.

For 75 % of companies applies the following: if the company does not manage the changes required to maximize the benefits (4.3) and service management does not work well (5.2) then the company is not growing. A company with the insufficient competency 4.3 is also not growing if either 5.5 (does not deploy new technology in the most effective way), or 5.3 together with 6.4 is missing (information asset management as well as asset and cost management is not good). Both rules hold for 75 % of SMEs.

4.4 Correlation between a Type of a Company and Its ICT Competencies

We also explored whether there was dependence between ICT competencies and a type of a company. In this experiment each company was first classified into one of three classes: ICT services (9 companies), services other than ICT (13 companies) and production (8 companies). Baseline accuracy if classifying all companies into the most frequent class) which actually means random classification, was 43.3 %. We used the same data as before. The highest overall accuracy over all the algorithms used was 72.4 % (Support Vector Machines). When compared with the baseline the accuracy was not too high. It implies a weak correlation between ICT competencies and the type of a company. Moreover, no correspondence was observed between long-term growth and a type of a company. From the ICT competencies it is the competency 2.5 (System and Process Innovation) that had the highest correlation with the type of a company. We also observed the influence of competency 5.3 but it was much weaker.

We also decided to join the class's services and ICT services but for this two-class problem – production vs. services – we received the same accuracy as for the classification into three classes. It supports our finding that there is no significant correspondence between ICT competencies and a type of a company.

A good insight into the correlation of a company type and its ICT competencies is displayed by association rules. The best rules, with a confidence higher than 0.9, are in the table below. '-' means the absence of a competency, 'X' means that the ICT competency holds. For instance '1.2 = - 11' means that absence of 1.2 competency holds for 11 companies. '2.5=X and Type=production 9' says that the presence of 2.5 competency together with the type of a company production holds for 9 companies.

1.2 = -11	=> Type = services 11	confidence = 1
1.2 = - and Type = services11	= > 2.5 = - 11	confidence = 1
1.2 = - and 2.5 = -11	=> Type = services 11	confidence = 1
1.2 = -11	=> 2.5 = - Type = services 11	confidence = 1
Typ = production 10	=> 1.2 = X 10	confidence = 1
2.5 = X and Type = production 9	=> 1.2 = X 9	confidence = 1
2.5 = - 15	=> Typ = services 14	confidence = 0.93
Typ = production 10	=> 2.5 = X 9	confidence = 0.9

5. Company Web Presentation

Web presentation is one of the most important marketing tools for establishing a potential relationship with business partners, suppliers and customers. From this point of view, mining a web presentation is a part of Marketing Intelligence [4], [6] because it helps in market decision making. In this and the following section the authors analyzed a relation between the web presentation and long term growth of a company. They were studying how web sites were designed and which information they offered. Authors analyzed what information on a web page was important for predicting the important web competencies in organizations. Some features – competencies – had to be confirmed by a human. However, the authors mostly focused on such features that could be extracted form web pages without human assistance.

A web presentation can be characterized by many factors. The most important ones are below.

- Company promotion (how the graphical design can catch the attention of the users).
- Product information (if the potential customer can find all necessary information about the product or service).
- Electronic transactions (if there is a possibility of online selling and payment).
- Web-enabling marketing (if there are any advantages for subsequent ordering, such as a loyalty program).
- Identifying of potential customers or partners (if there are possibilities of some business cooperation).
- Communication with customers and partners.
- Internal sharing of information and knowledge (if web site enables also internal communication of employees from outside).
- Other support services (such as after sale activities).
- Discussion with customers (if there is possibility of discussion with other customers their experience, ideas, problems, wishes).

5.1 Extracting Information from a Web Presentation

For examining a correspondence between the web presentation of a company and its long-

Tab. 3: Attributes of Web Presentations in Companies

Name	Description
Look	Evaluation by a human (1..5, 1=best, 5=worst)
Design	A method used for building the web presentation (tables, flash, cascade styles, other)
Language (6 attri.)	Language (LNG=Czech, English, German, Italian, Spanish, French) (yes/no)
Workpositions	Exists a page that offer a position in a company (yes/no)
Authentication	Authenticated area exists (yes/no)
Certificates	Certificate exists (yes/no)
References	List of customers and partners (yes/no)
News	A web page that contains news exists (yes/no)
Downloads	Information to download (exists/not exist)
Sitemap	Site map exists (yes/no)
E-order	To order a product (yes/no)
Search	Web search engine offered (yes/no)
ICQ	ICQ number (yes/no)
Menu.WORD	A set of binary attributes, one for each WORD that appeared in the menu
Text	Text that describes the profile of the company
Lemma	The same text extended with lemmata
PageRank	Google PageRank (numeric)
IndexedPage	Number of pages indexed by Google (numeric)
Backlinks	Number of backlinks to the web page (numeric)

Source: own

-term growth we focused especially on information that could be extracted automatically, i.e. without human assistance. The only exception is attribute Look which contains evaluation of a web presentation set by a human.

Attributes extracted from web presentations and used in this work are in Table 3 above. They can be split into four groups. First two attributes – Look, Design – concern the appearance. The second group contains features that characterize the contents of a presentation. The third group describes the text – in menu or in description of the profile of the company. The last group characterizes the importance of the web presentation in terms of web search engines Google and Yahoo. PageRank (from a link analysis algorithm used by Google) assigns a numerical weighting to each element of a hyperlinked set of documents with the purpose of measuring its relative importance within the set.

Backlinks are equal to the number of backlinks to the web page, obtained from Yahoo.

5.2 Methods

From 30 companies explored in the previous work [2] we removed two companies that had no web pages. For each company all the attributes from Table 3 were extracted. We also added information about ICT competencies described in Section 4 and the attribute that classified a company as growing or not growing. Companies were uniformly distributed to 14 companies with long-term growth and to 14 without long-term growth.

We also used various pre-processing techniques including natural language processing (NLP) tools for finding a small subset of important attributes and for ranking important words.

We employed two types of learning algorithms – classification and association mining. For finding associations we used Apriori algorithm [8], [24]. For classification we employed several learning algorithms – Naïve Bayes, Support Vector Machines, the instance-based learner kNN(N=3) and decision tree learner J48 [24]. As an evaluation criterion we chose again the overall accuracy, i.e. the ratio of correctly classified companies to the number of all companies. We used leave-one-out method for testing. All experiments were again performed with Weka package [24].

5.3 Correspondence between a Long-term Growth and Web Presentation

To find a relation between a long-term growth and the information on company web pages we performed three experiments. In the first experiment, when ignoring all text information, the decision tree learner J48 returned the best results and reached accuracy 82 %. It contained only two attributes, Backlinks (a number of backlinks of yahoo) and Workpositions (there is/is not a page that informs about job positions in the company).

In the next experiment we added words that appeared in menus and ignored the rest of the text. The accuracy did not increase (the highest accuracy was 75 %). The most significant attribute was again Backlinks and *reference* and *profile* words.

In the last experiments we used only text, i.e. the words that appeared on the web page, or text enriched with new features like bi-grams (two neighboring words) or lemmata. a lemma of nouns, adjectives etc. are a nominative in singular, a lemma of a verb is its infinitive form. However, new features did not increase the accuracy. It may have been caused by the fact that the web pages mostly contained only short texts. If the text consisted of longer sentences, it usually did not bring any relevant information. We also tried to use stoplist to remove nonsignificant words – prepositions, connectives or auxiliary verbs – but it did not improve the performance either. The best results were reached with Naïve Bayes classifier and Support Vector Machines (SMO). The highest accuracy augmented to 82.1 %.

When combining text with other attributes of web pages the accuracy decreased. Besides Backlinks and Workpositions mentioned above, among the most relevant attributes for long-term growing companies the principal role plays web design. Successful companies mostly used cascade styles. They also frequently offer product information to download, and a site map.

A strong positive correlation between PageRank and the long-term growth was observed. Also the higher PageRank and the higher probability are assumptions for the long-term growth.

5.4 Dependence between Web Presentation and a Type of a Company

Similarly as in the first part of this paper, the goal now was to learn whether there existed any dependence between the type of the company and other information, e.g. contents of the web pages. The best accuracy was not very high and reached 67.9 % when all information including the text of a web page was used. a decision tree (obtained with j48 algorithm, leave-one-out) is below:

```
not production (it means that in text there is no word "production")
|  not suppliers
|  |  not big
|  |  |  not new
|  |  |  |  Backlinks <= 2: ICT services (2.0)
|  |  |  |  Backlinks > 2: services (13.0)
|  |  |  new: ICT services (3.0)
|  |  big: ICT services (2.0/1.0)
|  suppliers: production (2.0)
production: production (6.0)
```

Tab. 4: Association Rules between Words and Other Information

Condition	Support	Conclusion==>	Support	Accuracy
already=1	8	Typ is not ICTservices	8	0.99226
German=1	7	Typ is not ICTservices	7	0.99157
Backlinks='(22-70.5]'	6	Typ is not ICTservices	6	0.99062
Backlinks='(70.5-186.5] '	5	Typ is not ICTservices	5	0.98929
Backlinks='(186.5-757.5] ' English mutation=0	5	Typ is not ICTservices	5	0.98929
Backlinks='(186.5-757.5] ' new=0	5	Typ is not ICTservices	5	0.98929

Source: own

We also performed experiments for a two--class problem. In this experiment we always took only one class, e.g. production, as positive examples and the rest, in this case ICT services and ICT services, as negative examples. For the class production, the accuracy significantly increased to 78.6 % even without using the text contents. After adding the text (all words that appeared at least 4x in the collection of all web pages) the accuracy augmented to 89.3 %. The most significant words for this class are the nouns producer, production which appear only here and never appear in the other classes. The other important word is provider. The same accuracy was obtained for ICT services vs. services + production even without the text on a web page. The most significant attribute was Backlinks and the most important words were the words new and information. For the class services the accuracy was very low – 75 % (Naive Bayes) if we compare this value with random classification (i.e. to the most frequent class) 71.4 %.

5.5 Association Rules

For finding association rules we employed Predictive Apriori [18], [24], which looks for the 'n' best rules. a rule is added to the output set if the expected predictive accuracy of this rule is among the 'n' best and it is not subsumed by a rule with at least the same expected predictive accuracy.

The most interesting rules concerned companies other than ICT services providers. Class association rules, i.e. rules that have a class in the conclusion, are in Tab. 4. a condition consists of words that appeared (Word=1) or not appeared (Word=0) on a web page. Support is a number of instances for which a condition (2nd column 2) and a conclusion (4th column) holds.

From the rules that combined words with other information extracted from web pages and had support greater than 5, the strongest rules was found for the companies from the class of non--ICT services.

Conclusion and Future Work

In this article we first explored the correspondence between long-term growth of SME and ICT competencies of their employees and then we examined the relation between a long-term growth and a web presentation. For these tasks we used various data mining algorithms – decision trees, cluster analysis, outlier detection and association rules. To our knowledge this is the first study that explores the relation between ICT competencies of a SME, a web presentation and a long-term growth.

In the first part we identified the most important ICT competencies that influenced a long term growth in the researched companies. We analyzed a correlation between ICT competencies and a long-term growth of a company. We showed that there is no unique com-

petency that would characterize growing or non-growing companies absolutely. It implies that it is always necessary to combine more competencies connected with business and information strategy. We employed an algorithm for learning a decision tree classifier that would classify growing and non-growing companies. The overall accuracy of the classifier was high and reached 83 %. The competencies which were connected with information strategy (2.1), ICT strategy alignment (2.2) business process design (2.3) and technology analysis resources (3.2) have the key importance. Then we analyzed the misclassified companies. We demonstrated that existing outlier detection algorithms could help to find the companies that were difficult to classify. Then we displayed the most interesting association rules that held for growing companies and for non-growing ones. To summarize, we found a small fraction of ICT competencies that were sufficient for recognition of the growing or non-growing company. It means that presence of business and information strategy support the key competencies from which it is possible to recognize growing companies. We also observed that there is no strong correspondence between ICT competencies and a type of a company.

In the second part we explored correlation between web presentation, ICT competencies and long term growth of a company. We proved that a long-term growing company could be recognized from the web presentation with 80 % and higher accuracy. We showed which attributes extracted from web presentation were the most important – communication with partners, English and German version of web pages, additional services offered to customers. We also found what information and what ICT competencies could be learned from the web presentation.

Although the set of SMEs was not too huge, this investigation shows how marketing intelligence can correspond with ICT and web competencies. It can help SMEs to improve their competitiveness, communication with customers and Internet marketing strategy. Some actual web pages of cooperation companies are now being designed with regards to the results of this study.

The findings of this study show that the company web presentation is not just commercial information on the Internet but, if it is well designed, it can improve communication with customers and suppliers. The proper marketing strategy supports the long term goals and can contribute to growth. It is also possible to predict from a web presentation whether the company puts emphasis on proper business strategy and is utilizes all the possibilities of ICT. These activities need employees with proper ICT competencies. Therefore the managers or owners of SMEs should support educational activities in this area and it is also a challenge for universities to help SMEs to increase their competitiveness with the educational internship. Also the importance of language mutations of web pages was proved.

In the near future economic results of the explored SMEs will be examined with ICT competencies and marketing changes. a company can be assumed to be successful if its offer meets the demand of customers. Unfortunately this information cannot be extracted from most of web presentations explored. The only exceptions are e-shops. Mining in customer transactions could answer this question. Future work will be also oriented towards the metrics of Google analytic methods and their contribution to improving marketing strategy and to the economic results.

References

[1] ANTLOVÁ, K. Motivation and Barriers of ICT Adoption in Small and Medium-Sized Enterprises. *E+M Ekonomie a Management.* 2009, Vol. 22, Iss. 2, pp. 140–154. ISSN 1212-3609.

[2] ANTLOVÁ, K., POPELÍNSKÝ, L. Predicting Information System Competencies in Small and Medium Enterprises with Data Mining Methods. In *Proceedings of IASK Intl. Conference E-Activity and Leading Technologies.* Madrid: IASK - International Association for the Scientific Knowledge, 2008, pp.133–139. ISBN 978-9-8995806-3-3.

[3] BARTON, G. J., LIVINGSTONE, C. D. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Computer Applications in the Biosciences.* 1993, Vol.9, Iss. 6, pp. 745–756. ISSN 0266-7061.

[4] BÜCHNER, A. G., MULVENNA, M. D. Discovering Internet marketing intelligence

through online analytical web usage mining. *ACM SIGMOD Record.* 1998, Vol. 27, Iss. 4, pp. 54–61. ISSN 0163-5808.

[5] CALDEIRA, M. M., WARD, J. M. Using Ressource-Based Theory to Interpret the Successful Adoption and Use of ICT in SME. *The 9th European Conference on IS.* Bled, Slovenia, June 27–29, 2001.

[6] CORNISH, S. L. Product Innovation and the Spatial Dynamics of Market Intelligence: Does Proximity to Markets Matter? *Economic Geography.* 1997, Vol. 73, Iss. 2, pp. 143–165. ISSN 1944-8287.

[7] FLACH, P., LACHICHE, N. IBC: a first-order Bayesian classifier. In DŽEROSKI S., FLACH, P. (eds.). *Proceedings of the 9th International, Workshop on Inductive Logic Programming.* Springer-Verlag. 1999, Vol. 1634. pp. 92–103.

[8] HAN, J., KAMBER, M. *Data Mining: Concepts and Techniques.* 2nd ed. Morgan Kaufmann, 2005. 800 p. ISBN 978-1558609013.

[9] HINCKLEY, S., PERL, V. The Competency of Global Manager. *Global Management.* 1996, Vol. 1, pp.138–148.

[10] LIU, B. *Web Data Mining.* 1st ed. Springer Verlag, 2008. 532 p. ISBN 3-540-37881-2.

[11] LEVY, M., POWELL, P. *Strategies for Growth in SMEs.* Oxford: Butterworth Heinemann, 2004. 420 p. ISBN 0-7506-6351-0.

[12] MCCLELLAND, D. C., BOYATZIS, R. E. Leadership motive pattern and long term success in management. *Journal of Applied Psychology.* 1982, Vol. 67, Iss. 6, pp. 737–743. ISSN 0021-9010.

[13] MITCHELL, T. *Machine Learning.* 1st ed. New York: McGraw Hill, 1997. 432 p. ISBN 0-07-042807-7.

[14] MITCHELL, T. *The Discipline of Machine Learning.* Pittsburgh, 2006. 7 p. TR CMU-ML-06-108, Carnegie Mellon University, Machine Learning Department. Available also on: <http://reports-archive.adm.cs.cmu.edu/anon/ml/CMU-ML-06-108.pdf>.

[15] POST, H. A. Building a Strategy on Competencies. *Long range Planning.* 1997, Vol. 30, Iss. 5, pp. 733–740. ISSN 0024-6301.

[16] PEPPARD, J., WARD, J. Beyond Strategic Information Systems: Towards an IS Capability. *The Journal of Strategic Information Systems.* 2004, Vol. 13, Iss. 2, pp.167–194. ISSN 0963-8687.

[17] SANCHEZ, D. Global Selling with Local Flavour. *Sales and Marketing Management.* 1998, Vol. 150, Iss. 8, pp. 26–28. Available also on: <http://www.allbusiness.com/marketing-advertising/4280237-1.html>.

[18] SCHEFFER, T. Finding Association Rules That Trade Support Optimally against Confidence. In DE RAEDT, L., SIEBES, A. (eds.). *Principles of Data Mining and Knowledge Discovery: 5th European Conference / PKDD 2001.* 1st ed. Springer, 2001. pp. 424–435. ISBN 3-540-42534-9.

[19] SCHMIEMANN, N. Enterprises by size class overview of SMEs in the EU. *Statistics in focus,* Luxembourg: Eurostat, 31/2008. 8 p. Available also on: <http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-SF-08-031/EN/KS-SF-08-031-EN.PDF>. ISSN 1977-0316.

[20] SPILIOPOULOU, M., POHLE, C., FAULSTICH, L. *Improving the effectiveness of a web site with web, Lecture Notes In Computer Science.* 1999, Vol. 1836. pp. 142–162. ISBN:3-540-67818-2.

[21] TAYLOR, T. Web Competencies for IT Students. In BRAUN, R. (ed.). *Proceedings 7th International Conference on Information Technology Based Higher Education and Training*, Sydney: University of Technology Sydney, 2006. pp. 563-570. ISBN 1-4244-0405-3.

[22] TORGO, L. Resource-bounded Fraud Detection. In NEVES et al. (eds.). *Progress in Artificial Intelligence: 13th Portuguese Conference on Artificial Intelligence, EPIA 2007.* Springer Verlag, 2007. 704 p. ISBN 3-540-77000-8.

[23] VESCOVI, T. Internet communication: the Italian SME case. *Corporate Communication: An International Journal.* 2000, Vol. 5, Iss. 2, pp. 107–112. ISSN 1356-3289.

[24] WITTEN I. H., FRANK, E. Data Mining. *Practical Machine Learning Tools and Techniques.* 2nd ed. Elsevier. 560 p. ISBN 0120884070.

[25] ZUFRYDEN, F., DRÈZE, X. Is Internet Advertising Ready For Prime Time? *Journal*

of Advertising Research. 1998, Vol. 38, Iss. 3, pp.7–18. ISSN 0021-8499.

[26] ZUFRYDEN, F., DRÈZE X. Testing Web Site Design and Promotional Content. *Journal of Advertising Research.* 1997, Vol. 37, Iss. 2, pp. 77–91. ISSN 0021-8499.

doc. Ing. Klára Antlová, Ph.D.
Technical University of Liberec
Faculty of Economics
Department of Informatics
klara.antlova@tul.cz

doc. RNDr. Lubomír Popelínský, Ph.D.
Masaryk University Brno
Faculty of Informatics
Department of Computer Science
popel@fi.muni.cz

Bc. Jindřich Tandler
Masaryk University Brno
Faculty of Informatics
173061@mail.muni.cz

Doručeno redakci: 7. 3. 2011
Recenzováno: 20. 5. 2011, 6. 6. 2011
Schváleno k publikování: 20. 9. 2011

ABSTRACT

LONG TERM GROWTH OF SME FROM THE VIEW OF ICT COMPETENCIES AND WEB PRESENTATIONS

Klára Antlová, Luboš Popelínský, Jindřich Tandler

How to increase competitiveness of SMEs? How to improve Internet marketing strategy? Which competencies of employees are the most important for the long term company growth? How to educate the employees? How to improve information and communication (ICT) using? The answers to these questions are examined in this article. In the last fifteen years the small and medium-sized companies have been searched by qualitative research. There have been defined seven main competencies which are connected with using ICT. The ICT competencies refer to a firm capacity to deploy their material and intangible resources, usually in combination with skills, to reach desired goals.These competencies have been associated with economic results and also with the companies´ web presentation. To solve this task, the authors used various data mining methods and web mining methods. The authors employed decision tree classifiers and association rules as well as class outlier detection methods and hierarchical cluster analysis.

The main goal of this contribution is to explore the correspondence between ICT competencies of SME, their web presentation, long term growth and to characterize a successful company. This can help an organization to improve employee's education and to develop all the competencies which are necessary for the successful implementation of ICT. During the long term research the data have been collected within thirty enterprises using multiple interviews. Some companies had developed significantly more ICT competencies than the others and these competencies help them to be more successful and competitive.

Key Words: *information and communication strategy, small and medium-sized companies, data mining, web mining, marketing intelligence, competencies.*

JEL Classification: *C81.*