INTRODUCTION TO UNSTRUCTURED DATA IN INNOVATION PROCESS MANAGEMENT

Lodz University of Technology

Prepared by: Piotr Józwiak, Iwona Staniec
Date: October 2015
Draft Version 1/1
Estimated number of lectures/teaching sessions needed:

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

INTRODUCTION

Introduction to unstructured data in innovation process management is very important to the present and future organizations. Unstructured data is a generic term used to describe data that doesn't fit in column row databases and is a mixture of text and non text data. The challenge for businesses is to develop processes to apply structure to the unstructured. For example determining the level of satisfaction of customers by analyzing emails and social media may involve searching for words or phrases. Words and phrases may be grouped into positive, negative or neutral classifications. Opportunity lies in understanding how adding unstructured data to the mix creates competitive advantage.

PART 1 Innovation process

According to Schumpeter [Schupeter K. (1912) The Theory of Economic Development: AN Inquiry into Profits, Capital, Credit, Interest and the Business Cycle, Cambridge: Harvard University press, p66] We can distinguish five types of innovation:

1. The introduction of a new good or service,
2. The introduction of a new method of production,
3. Opening of a new market,
4. Conquest of a new source of supply of raw materials of half manufactured goods,
5. Implementationof a new form of organization.

   It should be noted that in his aproach innovation is always related to introduction to market or any other form of profiting from implemenation of new solutions to organization.

And this list still covers most of situation and is used for innovation activities. Some approaches has changed as distinguishing in Oslo Manual:

- product innovation,
- process innovation,
- marketing innovation
- organisational innovation. [Oslo Manual OECD http://www.oecd.org/sti/inno/2367580.pdf]

You should also notice trends in opening the organizion to markets with Open Innovation approach proposed by Henry Chesbrough [Research-Technology Management • July—August 2012]. Where in his open innovation model [Research-Technology Management • July—August 2012 p.23]. he focuses on inputs coming from not only internal technology base but also from external sources. During the

process of development ideas not selected for market introduction should not go to company archives but should be offered to market in form of licensing or spin-offs creation. This strategy is mostly adopted in high tech industries.

Analyzing innovation activities performance regardless of models and approaches we should identify performance indicators. List of such is presented in table 1.

Table 1. Innovation KPI's

Task	**KPI's**
Start-up Costs	• The number of full time staff involved • Operating expenses • Capital expenditure
Speed	• Actual time to market • Time to key checkpoints • Actual versus planned full-time employee hours
Scale	• Actual versus planned volume produced • Actual versus planned product availability • Actual versus planned first year sales • Actual versus planned distribution • Actual versus planned timing of ad campaigns
Support Costs	• The extent to which new offerings cannibalize existing products • Marketing and promotional activities • Pricing actions • Key staff devoted to the project • Product maintenance and service cost

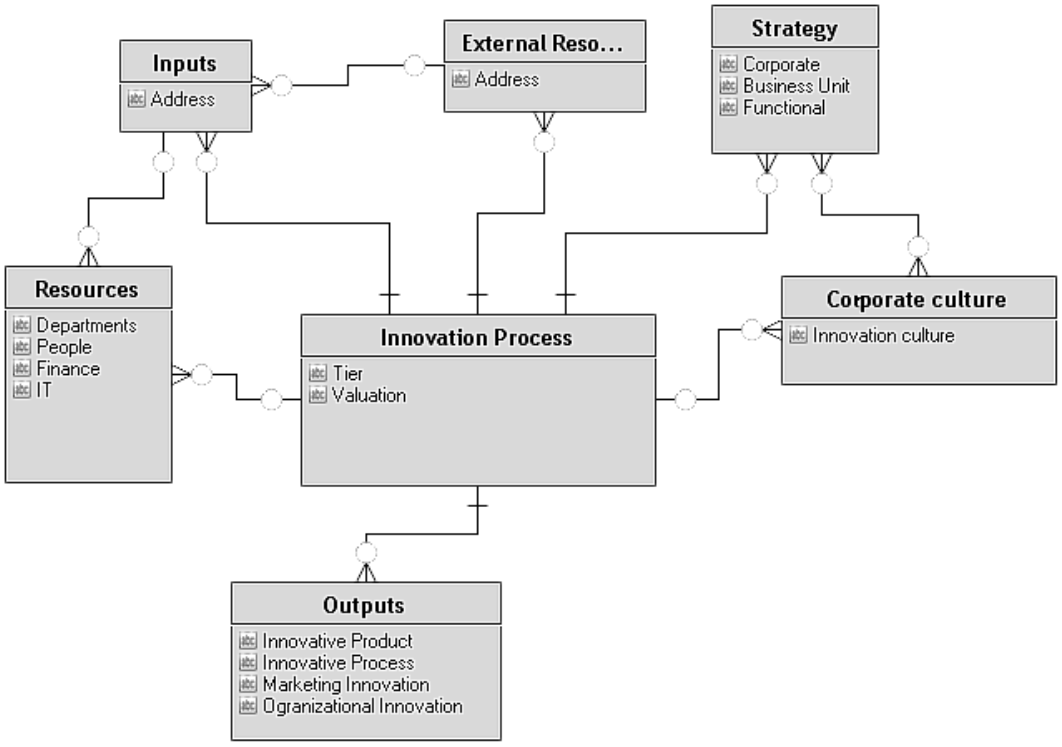
Source: Abhishek Gupta, A Study of Metrics and Measures to Measure Innovation at Firm Level & at National Level, June 2009 http://www.imri.dauphine.fr/fileadmin/mediatheque/centres/IMRI/2009-03.pdf p.22

Presented KPI's cover various stages of innovative process and cover metrics from simple financial as operating expenses through based on time as time to market to commercialization stage as indicators for marketing and promotional activities. The list is not comprehensive and also other measures can be considered and in case of innovation most often used in companies analyzed by Boston Consulting Group in years 2007-2010 were:

- customer satisfaction,
- overall revenue growth,
- margins,
- return on innovation spending,
- total funds invest in growth projects,
- revenue realized from offerings launched in the past three years,
- allocation of investments across projects,
- projected versus actual performance,
- percentage of sales from new products,
- new product success ratio,
- number of new product services,
- time to market,
- projected versus actual performance,
- average development time per project
- number of projects that meet planned targets,
- percentage of ideas funded,
- number of ideas killed or table at each milestone,

- cannibalization of existing product sales by new offerings,
- patents. [1,2,3]

In case of BCG research most list was created from those which were repeated in years 2007 to 2010. Interesting indicators cover are of cannibalization of existing products and revenues from new product for three years.



Drawing 1. Innovation entity relationship model

Source: Own research

In presented innovation entity model process is dependent on tiers and valuation such representation gives flexibility to adapt to various approaches and methodologies which could be implemented in system as Stage Gate or project management methodologies as Prince 2,

IPMA, Scrum and other. Sample Tiers classification Where going from Innovation Process we can classify Tiers and Valuation and Tiers can be classified as Scoping, Idea Generation, Development, Testing, Lunch.

Tasks

- Tasks for students define innovation activites in organization
- Identify data sources for innovation in organization you know (or think about startup)

Questions

- What is innovation
- Which stages of innovation?
- What is data source?

Literature:

1. J. Andrew, J. Manget, D. C. Michael, A. Taylor, Innovation 2010 A Return to Prominence-and Emergence of New World Order, April 2010, BCG, http://www.bcg.com

2. J. Andrew, K. Haanæs, D. Michael, H. L. Sirkin, A. Taylor, Measuring Innovation 2009: The Need for Action, April 2009, BCG, http://www.bcg.com

3. J. Andrew, H. L. Sirkin, K. Haanæs, D. C. Michael, Measuring Innovation 2007 A BCG Senior Management Survey AUGUST 01, 2007

4. P. Miller, K. Klokgieters, A. Brankovic, F. Duppen, Innovation leadership study, Managing innovation: an insider perspective, Capgemini Consulting, IESE Business School, University of Navarra, march 2012

PART 2 Unstructured data

In case of companies there is many various sources of data in both structured (for example in tables like invoices) and unstructures (for example – customer emails or reports) as transactional data, social media discussions, product reviews, email, claims, internal reports, intranet, forums, archives, public data.
As those data are vast in many companies and coming from various sources new aproches as BigData appeared on market to cope with such data diversity and analysis solutions. Also algorithms for analysis were actively developed in last years as machine learning, artificial intelligence (a subject within computer science), discipline concerned with the implementation of computer software that can learn autonomously.[http://www.britannica.com/technology/machine-learning]
Expert systems and data mining programs are the most common applications for improving algorithms through the use of machine learning. Among the most common approaches are the use of artificial neural networks (weighted decision paths) and genetic algorithms (symbols "bred" and culled by algorithms to produce successively fitter programs). [http://www.britannica.com/technology/machine-learning]

- The first is a grouping of algorithms by the learning style.
- The second is a grouping of algorithms by similarity in form or function (like grouping similar animals together). [http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/]

For performing analyses and algorithms development there is many software packages for data analyses. Selected open source products are presented in table below.

Characteristic	RapidMiner	R	Weka	Orange	KNIME	scikit-learn
Developer:	RapidMiner, Germany	worldwide development	Univ. of Waikato, New Zealand	Univ. of Ljubljana, Slovenia	KNIME.com AG,Switzerland	multiple; support: INRIA, Google
Programming language:	Java	C, Fortran, R	Java	C++, Python, Qt framew.	Java	Python+NumPy+ SciPy+matplotlib
License:	open s. (v.5 or lower); closed s., free Starter ed. (v.6)	free software, GNU GPL 2+	open source, GNU GPL 3	open source, GNU GPL 3	open source, GNU GPL 3	FreeBSD
Current version:	6	3.02	3.6.10	2.7	2.9.1	0.14.1
GUI / command line:	GUI	both; (GUI for DM = Rattle)	both	both	GUI	command line
Main purpose:	general data mining	sci. computation and statistics	general data mining	general data mining	general data mining	machine learning package add-on
Community support (est.):	large (~200 000 users)	very large (~ 2 M users)	large	moderate	moderate (~ 15 000 users)	moderate

Image: Open source Data mininig tools comparison
http://www.infoivy.com/2014/06/not-all-data-mining-packages-are.html

There is also market for paid closed licence solutions from companies as ORACLE, SAS, IBM SPSS, STATISTICA and also software libraries form Java, Python and other programming languages.

Data analysis process

Data analysis procedures should start from stating research question (hypotheses) which should be validated during research. Research hypotheses could be stated on various stages of innovation process. In many cases main interest would be related to actual or new product development but could also be related to strategies, inputs and outputs. With next step measurements indicators should be set up to allow for analysis results assessments. For this benchmarking methods could be used or KPI set on research supported tasks. Then the process of actual analysis as on next image can be setup. And then implemented in software.

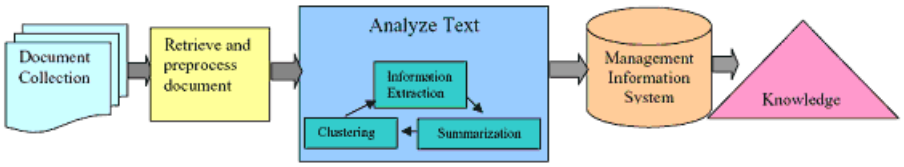


Image: Text mining process [http://www.mu-sigma.com/analytics/thought_leadership/cafe-cerebral-text-mining.html]

For analysis there should be relevant data collected as patents, articles, news feeds, internal and external reports, customer based data. Data collection could be most time and resources consuming tasks. But in the process tools for scraping data from internet sources can be used. Such solutions are integrated in some data analysis packages (as rapidminer) and could bu used in many cases to capture data from well structured websites. But in other cases there would be necessisty to write special pieces of code to populate database with relevant information or use external software as IFTTT which allows to automate various online task (or other paid tools).

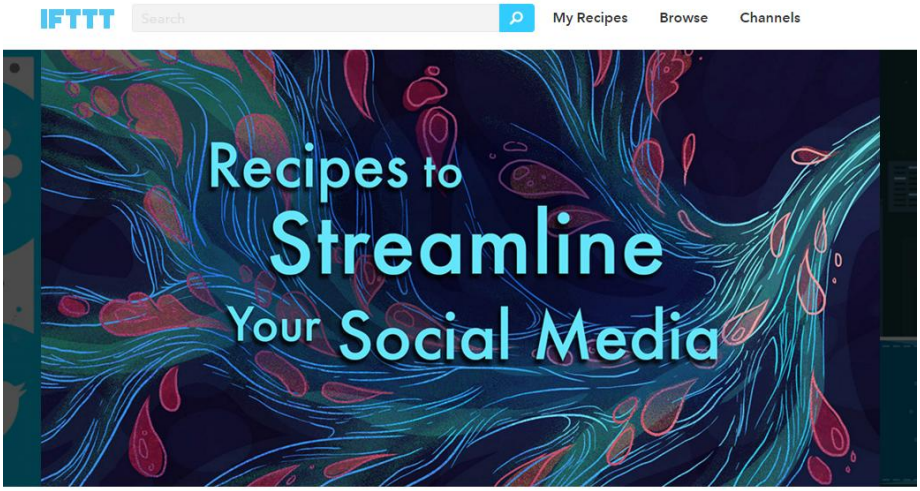


Image: IFTTT: service screenshot with premade task for social media.
Source: ift.com screenshot

Next step would be to clean the data from unnecessary information. In case of website data scraping in some cases there would be a need to remove some HTML code, or unnecessary information as website name, dates etc. In case of text analyses this task could be even more important and time consuming as often repeated keywords should be removed and to spot this at leas partial analysis should be performed. Such sample analysis with where data pre-processing could be included is presented on next image.

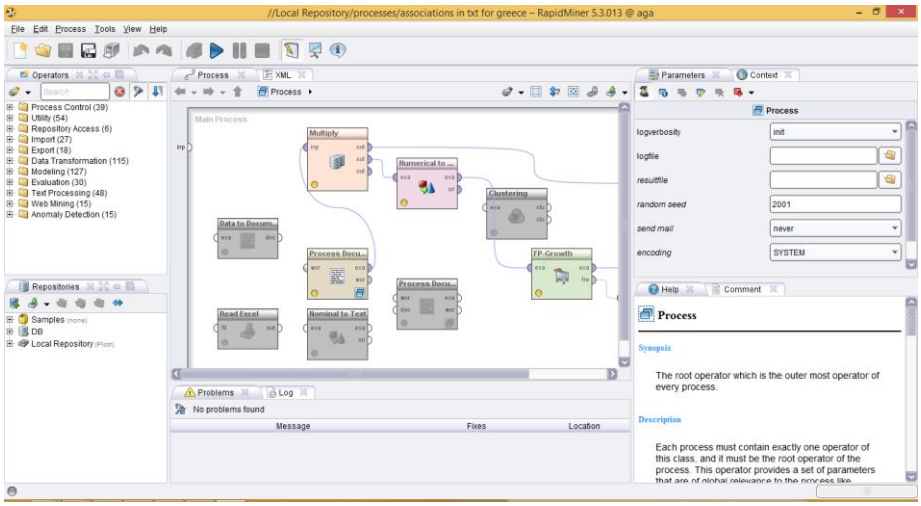


Image: Sample implementation of process in Rapidminer
Source: Rapidminer screenshot

Then results can be presented in various forms as tables, chards or graphs as in next image.

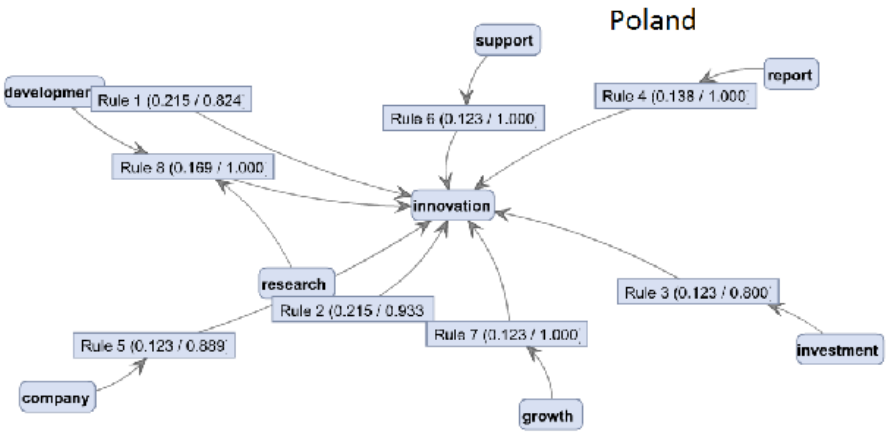


Image: Proces results text-mining analyses ISOM graph
Source: own research

Tasks:

- Select innovation related subtask in company and give information on data sources which would support decision making on this step.
- Setup automated data acquisition process in software you decide?
- Import gathered data into software of your choice.

Questions:

- How important is data cleaning in data analysis process?
- What are the forms of analysis results presentation
- Which data sources companies can use?

PART 3 ANALYSIS OF UNSTRUCTURED DATA

Should unstructured data be analysed?

As J. Kelly[1] said, *"If you do not use data and information to make better decisions and increase your efficiency, you will stay behind. This applies to businesses, governments, hospitals and indeed any organization. That is why, it is so important."* The concept of unstructured data does not mean that it does not have any structure, but this is data that does not fit into any database, nor can it be part of the relational data model in the company. We try to arrange it somehow to provide analysis, but the problem is that it does not arrive on time to employees who need it to be productive. Often, in the theory of decision it is said about such situations that we are dealing with incomplete information[2]. Once, until the 90s, it meant that there was no such information, and today - that we did not get to it in time. That is why it is very important to pass on relevant information to the right people at the right time and in the appropriate format. According to the Gartner[3], search for documents by professional staff takes 10-50% of their working time, and reading documents in order to obtain from them the necessary data is another 5-15% of their working time. Delving further into the statistics[4], we see that on average about 80% of the data in companies considered to be unstructured, and over half of the latest information technology purchases of companies in 2014-2015 was related to information management systems (knowledge) including decentralized model, under which it is possible to share data and perform own analyses. This is also confirmed by the producers of solutions of ECM class(Enterprise Content Management) , who emphasize in their reports that only about 20-25% of data in organizations is structured and can be managed by using the current tools (ERP, CRM, SCM, PLM, etc.)[5]. According to the MIT Centre for Digital Business, the companies that will operate on the basis of unstructured data and use not only structural data in decision-making

[1] J. Kelly *Big Data: From Promise to Reality*, Analysis Big Data 2015.
[2] M. Weber, *Decision making with incomplete information*, European Journal of Operational Research Volume 28, Issue 1, January 1987, pp.44–57.
[3] http://www.gartner.com/technology/topics/trends.jsp online 10.10.2015.
[4] Gartner analysts' report: 2015 Magic Quadrant for Business Intelligence and Analytics Platforms.
[5] S. Malecki, *Dokument, treść, informacja, wiedza*, Outsourcing Magazine, 2006.

will enjoy the productivity and performance 5-6% higher than it would be expected based on other investments and used technology.
Research[6] shows that the companies that want to better understand the processes anticipate upcoming events respond to the threat and use opportunistic chances need to improve information systems through the use of unstructured data. It is also forced by the turbulence of the environment through project work requiring dispersion of information and increasing the speed of their movement within the corporate network. In addition, research conducted by the EMC Forum 2013[7] indicates that:

- 39% of entrepreneurs believe that unstructured data management ensures the success of the company,

• 19% of entrepreneurs believe that by managing unstructured data has reached a competitive advantage,
• 36% of entrepreneurs believe that the introduction of unstructured data management increased the safety and security the data available.

What does it mean?

In modern enterprises, there is a large number of unstructured data that are included in text documents, presentations, graphics, emails, audio and video files, websites and all sorts of office software. In addition, part of business processes is in the form of unstructured data in the minds of employees. Information value of unstructured data is huge and unfortunately difficult to specific estimate. How this information is important is confirmed if we attempt to imagine what would happen if we do not have it? Unfortunately, negligence in this area in a crisis situation could result in spectacular mistakes of people in different positions. On the other hand, neatness in the information and accurate search translate to acceleration and improvement of the quality of business processes. As a consequence, it has a great impact on competitive advantage in an ever more dynamic environment, e.g. for the Fukushima the consequences were devastating, as Martin White[8] put the thesis that "*the eruption of a nuclear power plant in Fukushima occurred due to a failure to find in time the current technical documentation describing the procedures for reactor shutdown. Just after the tsunami, there was very little time to undertake this action and in an emergency situation the staff failed to perform this*

[6] I. Staniec *Uwarunkowania skuteczności zarządzania ryzykiem w organizacjach*, Politechnika Łódzka 2011, pp. 149-171.
[7] http://poland.emc.com/infographics/succeed-big-data.htm online 10.10.2015.
[8] Nee C., White M., Woolford K., Pascu T., Barker L., Wainwright L. (2015) *New methods for examining expertise in burglars in natural and simulated environments: preliminary findings.* Psychology, Crime & Law, 21 (5). pp. 507-513.

procedure because no one could quickly find out how to do it correctly. "

Another new source of unstructured data is information from all kinds of social media (e.g. Facebook, Nk.pl, Google + Twitter, Blip) and Internet resources. It is difficult to analyze, because it rarely contains specific numerical values, and it changes over time, but can be tested for the presence of keywords, frequency of entries or the use of selected words or phrases, appearance of positive or negative connotations, and so on. By analyzing this data, a fuller picture of the expectations and needs is obtained, e.g. it is possible to: identify customers, adjust the offer to the customer, check market reaction to the product, price the product to others, compare the response to competitive products or evaluate the risks of business decisions.

How to analyze them?

These data cannot be analyzed by means of traditional tools. Firstly, because there are a lot of them and they change dynamically. Secondly, qualitative techniques, including automatic text analysis methods, are still underdeveloped - it is necessary to pre-structure the collected information for text analyses. Thirdly, presently we face a significant shortage of analytical and managerial talents necessary for the proper use of integrated structured and unstructured data[9]. Therefore, more and more IT professionals and business users closely monitor the solutions that could cope with this mass of knowledge and data.

An important change in the analytical approach used to date is the introduction of intelligence and automation to the process of digitisation[10] of unstructured data. This means that documents are classified according to their content, i.e. information is captured contextually, and then validated and transmitted smoothly, e.g. into core business applications and document flow processes. Smart digitizing with a high level of accuracy can instantly change the data from unstructured to structured[1111]. As such, they become part of the data model that can be indexed and integrated. Only with this approach

---

[9] Experts from SAS claim that there are too few relevant specialists (particularly mathematicians). McKinsey Global Institute Report (http://www.mckinsey.com/insights/business_technology/open_data_unlocking _innovation_and_performance_with _liquid_information online 10.10.2015) shows that in 2015 in the USA itself, there is a shortage of over 140,000 workers with the skills that allow for the use of structured and unstructured data and up to 1.5 million managers and analysts needed for data analysis and decision-making.

[10] D. Larose Ch. Larose *Data Mining and Predictive Analytics* Wiley, 2015.

[11] Kantardzic M., *Data mining: concepts, models, methods, and algorithms*, Wiley-Interscience: IEEE Press, 2003.

achieving added value of this content is easier and it is possible to search, preview and analyze the content within particular organizational units.
The market offers you a variety of solutions; also their basic standards have stabilized allowing for a relatively reasonable planning so in the realm of technology, as well as in understanding of the Total Cost of Ownership (TCO). Increasingly, there appear also services such as outsourcing e.g. of document management – ArchiDoc[12]. While establishing systems for digitisation of unstructured data, attention is paid to the roles, procedures and assignment of information.

Information	**Processes**	**Organization**
Catalogue of business-relevant data used by the organization. Allows for introduction of the same description language for all	Uniform information flow processes. They govern the handling of data	It defines the role played by individual organizational units in the processes and what information is necessary for them

And in further proceedings, it is necessary to determine the approach to unstructured data through:
1. Search for information
2. The use of information.
From the point of view of data analysis, important is:
1. The data collection as part of an established data management system by converting the recorded data and certain accesses.
2. The preliminary analysis consisting in discovering the necessary information by analyzing the context, text mining, extraction of concepts and so on.
3. Organizing the data by determining categorization, ontology[13], taxonomy[14], abstracts, or brief descriptions of the content, and

[12] According to a survey by Ipsos, in 2014 outsourcing in this area was used by about 25% of companies, predominantly law firms and companies from the SME sector.
[13] Ontology - the formal representation of a field of knowledge, which consists of the record of collections of concepts (English: concept) and the relationship between them. This record creates a conceptual scheme which, being a description of a given field of knowledge, can also serve as a basis for requesting the properties described by ontology of concepts. In computing, it has been operating since 1967, but gained significance in the 1990s thanks to Thomas Gruber [T. R. Gruber. *A translation*

principles of deduplication. Very important at this stage is the organization of work with the data because that is what allows for further use and quick access to information.
Therefore, further steps to organize unstructured data include:

- analysis and visualization,
- statistics of used terms, identification of similar documents,
- graphical presentation of the relationship between the data already structured,
- automatic classification, taxonomy involving the organization of topics in hierarchies and grouping into default or pre-defined classes
- predictive modelling, using only text, e.g. predicting the customer's attitude based on his comments; or text and other data, e.g. predicting future purchases based on the opinions and demographic information.

Pre-treatment of the information contained in the text involves identification of the text units: paragraphs, expressions, words, phrases, etc. It is important to introduce the so-called. stop list, or exclude irrelevant words and phrases that often occur, but are useless in the analysis because they do not convey any meaning. Unfortunately, from the point of view of the accuracy of the analysis it is necessary to bring the words to their basic grammatical form, their standardization, and the use of synonyms by analyzing text collection. The data obtained in this way is the basis for further analyses such as data mining connected with finding information, search for patterns, clustering of documents, automatic generation of abstracts or keywords.
Practices that help in the analysis of unstructured data are, for example, the construction of the subject information model, the vocabulary or corporate taxonomy, classification of the content used in the company, coherent system of content tagging, storage and retrieval of information. These are not complicated things nor are they long in implementation. It is rather the process of organizing, which should be started and continued alongside the changing information environment. It is recommended that a RCFA analysis (Root Cause Failure Analysis) should be conducted before the introduction of the information seeking process. Since this method is focused on identifying the sources of problems or incidents so it allows for adapting into the organizational culture nomenclature of groups (classes), namely the introduction of appropriate ontologies and

---

approach to portable ontology specifications. „Knowledge Acquisition". 5 (2), pp. 199–220, 1993. DOI: 10.1006/knac.1993.1008]

[14] Taxonomy is a system, division into groups according to a certain system of objects or concepts.

taxonomies to facilitate data retrieval. In this case, the root cause is identified as impossibility to get on time to the required information. Identification and removal of the root causes by focusing on correction prevents the occurrence of such problems in the future. In this case RCFA is used as an interactive process and a tool for continuous improvement of automatic digitalisation of the data collected.
The basis for automation are the tools of artificial intelligence[15], dealing with creating of intelligent behaviour models and programs simulating this behaviour.
Use of information aims to power analytical and reporting systems with the relevant data, set alerts, or early warning signals, perform predictive analytics, utilize, index and search, and conduct operational reporting .
As in any field, in the case of combined structured and unstructured data, analyses start with a clear definition of the purpose and the data sources used.

Table 1. Examples of application areas, objectives and data sources specified for them

Area	Objectives of analysis	Data sources
customer relationship management	improving the quality of services	e-mails, complaints, opinions, call centre records, proposals to improve, research reports
compliance with the law	detection of irregularities	Financial reports, financial news, company documents, commercial records, e-mails, opinions of employees, co-workers and customers

Then in the next step, it is expanded by the concrete ranges of analysis, which are worth carrying out.

[15] Genesereth M., Nilsson N., Logical Foundations of Artificial Intelligence, Morgan Kaufmann Inc., 1989.

Table 2. Examples of areas of analyses conducted for the area of customer relationship management depending on the purpose.

Objectives of the analysis	Improving the quality of services	Improving the process of customer retention through early detection of the risk of leaving
Data sources	e-mails, complaints, opinions, call centre records, proposals to improve, research reports	e-mails, complaints, opinions, call centre records, proposals to improve, research reports, posts on the forums
Analysis range	1. Analysis of the phrases most often cited in customer statements relating to service 2. Analysis of the most frequently repeated words compounds in the context of a given service leading to the development of a map of concepts most frequently associated with the service 3. Analysis of comments on the given service in terms of its features and attributes 4. Analysis of the use in the statement of positive, negative or synonymous expressions 5. Comparative analysis of the statement on the competitor's service	1. Analysis of the level of customer satisfaction based on a defined index[16] 2. Monitoring changes in an index for the selected customer in time 3. The definition of alarm thresholds with changes in the index value, meaning an increased risk of loss of customer 4. Identification of the causes that led to a potential reduction of the index value for the chosen customer 5. Identification of communication patterns and their occurrence in time, prior to the customer's decision to abandon the services
Variables used	Time Customer segment	Time Chosen customer

Examples of advanced IT tools for exploration of unstructured data:

[16] Indices (indices of measurement) are defined depending on the needs of the entity.

1. **Teradata Aster Discovery Platform**[17] - is an analytic platform from the area of big data[18] that has the ability to acquire and analyze data of any format and from a variety of sources. These can be structured and unstructured data: plain text, billing data, data from the Internet, any multi-structured data. Teradata Aster Discovery Platform can actually process the data and, using analytical tools, find the most valuable ones. It also allows for quick testing of business hypotheses and presenting their results in a user-friendly visualization environment. The platform is available in the offer model, the so called services in the cloud or a traditional model of deployment of solutions at the client's.
2. **Unified Data Architecture**[19] - is one of the best and most complete solutions available on the market for advanced business analytics. The result of the interconnection of databases of Teradata Aster Discovery Platform and open-sourcing Hadoop platform is a unified, high-performance analytical environment for the enterprise. Organizations can ask here questions about a broad spectrum of analyzes carried out on the basis of any type of data, at any time, and discover new and valuable dependencies, which will result in higher productivity, lower costs and new business opportunities.
3. **Automatic Business Modeler (IBM)**[20] maps the business issues onto Machine Learning algorithms and their parameters and settings, processes optimization algorithms, and then presents the results in the form of business solutions. ABM also allows full automation of the necessary but time-consuming tasks associated with construction of predictive models, such as the selection of variables for analysis, transformation of variables or choosing the best model. Available online, in SaaS model (Software-as-a-Service).
4. **Hadoop**[21] is a popular modern open platform that allows you to collect and analyze large sets of data from sources such as social networking, history of visits to websites, logon servers,

---

[17] http://www.nimbusninety.com/sites/default/files/documents/teradata/wp-TDAster-Discovery-Platform.pdf online 27.10.2015.

[18] *Big date* – a term for a large variety of variables and data sets whose processing and analysis is difficult but also valuable because it may lead to gaining new knowledge [*Lee J.; Lapira E.; Bagheri B.; Hung-An K. (2013). Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment. Manufacturing Letters **1** (1): 38–41. doi:10.1016/j.mfglet.2013.09.005]*.

[19] http://www.thebigdatainsightgroup.com/site/sites/default/files/Teradata%20Whitepaper_0.pdf online 27.10.2015.

[20] http://algolytics.com/products/automatic-business-modeler/ online 27.01.2015.

[21] http://hortonworks.com/use-cases/sensor-data-hadoop-example/ online 27.01.2015.

transaction systems, video files collections, or sensory data of related devices.

Task for students

1. Define the exemplary ranges of analysis and variables used for the area of compliance with the law and the objective shown in Table 1.
2. Define exemplary objectives, ranges of analyzes, data sources and variables used for the area of human resources management.
3. Define the exemplary objectives, ranges of analyzes, data sources and variables used for the area of sales management.
4. Define the exemplary objectives, ranges of analyzes, data sources and variables used for the area of public relations.

Questions for students

1. Why to analyze unstructured data?
2. What are the practical applications of RCFA analysis?
3. What is the information value of unstructured data?
4. What are the stages of searching for information in unstructured data?

REFERENCES

Analyst report Gartner: 2015 Magic Quadrant for Business Intelligence and Analytics Platforms.

Genesereth M., Nilsson N., (1989) *Logical Foundations of Artificial Intelligence*, Morgan Kaufmann Inc.

Gruber T. R. (1993) *A translation approach to portable ontology specifications*. „Knowledge Acquisition". 5 (2), s. 199–220. doi: 10.1006/knac.1993.1008

http://www.nimbusninety.com/sites/default/files/documents/teradata/wp-TDAster-Discovery-Platform.pdf online 27.10.2015

Kantardzic M., (2003) *Data mining: concepts, models, methods, and algorithms*, Wiley-Interscience: IEEE Press,.

Kelly J. (2015) *Big Data: From Promise to Realityby*, Analysis Big Data.

Larose D. Larose Ch. (2015) *Data Mining and Predictive Analytics* Wiley.

Lee J.; Lapira E.; Bagheri B.; Hung-An K. *(2013). Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment.* Manufacturing Letters **1** (1): 38–41. doi:10.1016/j.mfglet.2013.09.005.

Malecki S., (2006) *Dokument, treść, informacja, wiedza*, Outsourcing Magazine.

Nee C., White M., Woolford K., Pascu T., Barker L., Wainwright L. (2015) *New methods for examining expertise in burglars in natural and simulated environments: preliminary findings.* Psychology, Crime & Law, 21 (5). pp. 507-513

Report McKinsey Global Institute http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information online 10.10.2015

Staniec I. (2011) *Uwarunkowania skuteczności zarządzania ryzykiem w organizacjach*, Politechnika Łódzka.

Weber M., (1987) *Decision making with incomplete information*, European Journal of Operational Research Volume 28, Issue 1, January 1987, pp. 44–57.

Other references

http://www.thebigdatainsightgroup.com/site/sites/default/files/Teradata%20Whitepaper_0.pdf online 27.10.2015

http://algolytics.com/products/automatic-business-modeler/ online 27.01.2015

http://hortonworks.com/use-cases/sensor-data-hadoop-example/ online 27.01.2015

http://www.gartner.com/technology/topics/trends.jsp online 10.10.2015.

http://poland.emc.com/infographics/succeed-big-data.htm online 10.10.2015.