Eye Tracking Data Acquisition and Preprocessing for Modeling of Manager Knowledge in Financial Report Analysis

Adrian Kaźmierczak
Wrocław University of Economics, Komandorska Str. 118/120, 53-345 Wroclaw, Poland
Email: adrian.kazmierczak@ue.wroc.pl

Abstract—**The article presents process of data acquisition and detection of pattern in the way of perceiving economic data. For this purpose, results of eye tracking study were used to obtain information on how data are analyzed by SME managers in accordance with their level of economic and financial knowledge. As a result, information was derived on viewing patterns allowing for generation of sequence patterns. Research results will help design a mechanism for intelligent interpretation of economic information in Business Intelligence systems.**

I. Introduction

ANALYSIS of a manager's eye movements during the time of reading financial reports constitutes one of the new areas of research. The objective is to identify perception processes, both overt and covert, for a participant's consciousness. Making eye movements in analysis and decision systems, as studies have shown [1], can play an active role in decision-making [2]. Eye tracking allows for very precise time measurement of perceiving data, for carrying out an operation and reaching a decision, which would not have been possible with other traditional methods.

Studies show that a significant proportion of managers of small and medium-sized enterprises (SMEs) have no background knowledge and qualification in economics making it difficult for them, on the one hand, to conduct correct data analysis and financial reports and, on the other, to make complete use of the many features and services available in decision support systems [3]. Universal, "made-to-measure" solutions prevent adaptation of system interface to the level of knowledge, skills and qualifications of a SMEs manager. Effectiveness of decision-making is impaired by lack of financial analysts in SMEs as well as lack of sufficient funds to cover the high costs of continuous business consultation. Their decisions, managers in SMEs need systemic support based on data from transactional systems. By processing these data, applying financial analysis tools help in the preparation of decision support projects. Advancement in information technology has provided an opportunity for managers to access diversified information stored in various databases and Internet resources. Excessive amounts of report data usually generated by transactional and dedicated decision support systems, among others, constitute a problem. When it comes to business processes, information overload significantly reduces one's ability to reach the right decision, in particular in the case of SMEs.

The aim of the article is to show the possibility of preparing data for patterns of sequence of eye-glance, or patterns of manager's analytical thinking using eye tracking system. Due to inability to automate process of generating data, it was necessary to find a solution that would make it possible to prepare them beforehand. The article will describe method of data processing, in the context of conducted research. These data could be assisted built ontologies relevant to managers in the knowledge areas of economic and financial, understood as formal conceptual layer specifications [2].

The structure of the article is as follows. The next section of the article describes the study that led to the research data. Another point is devoted to the possibility of generating a sequence of looking at financial reports. Taking into account, in point 4, the description of how to obtain sequence patterns achieved using data received. The article concludes with a summary description of experiments, showing the scope of planned research.

II. Analysis Of Financial Statement

Research has been conducted on interpretation of financial statements with the aim of modeling of a manager's knowledge. The purpose of the experiment was to examine perceptions of participants of different levels of knowledge in the field of financial reporting. Experienced managers as well as beginner analysts - economics students, took part in the study. During the experiment, attempt was made to acquire information on perception of report through the eyes of participants, as well as expert's and novice's patterns of observations. The purpose of this study was to use this information in modeling a manager's knowledge profile, specifying sequences and relationships in how appearance of presented reports is perceived.

The study, assumed that a manager uses business-oriented software (such as DSS/ERP/BI) supported by financial ontology, analytical models, and appropriate expertise.

In the experiment, participants were asked to evaluate an enterprise's economic and financial situation. Case study consisted of financial reports and adequate ontology prepared and made available to the participants.

The experiment started with an analysis of an internal managerial report containing widely used indicators of assessing financial situation, namely: liquidity and

profitability, debt, and turnover ratios. The most important factor directly observed by participants in the experiment was a significant increase in profitability in an analyzed period.

An ontological approach for conceptualization of knowledge domain models was adopted in the study. Development of ontology for selected areas of economic and financial knowledge important to SME manager constitutes an important element. Ontologies are created by experts in the area, with the consideration of acquisition of hidden knowledge as a result of research using eye tracking, for example in order to correctly explain profitability ratio, it was necessary to make use of ontology describing the concept of profitability, as presented in Figure 1.
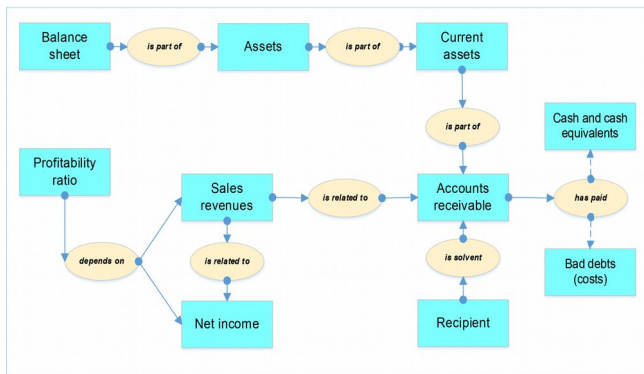

Fig. 1. Ontology: view of profitability

Many research projects show that an ontology of economic and financial indicators aid in decision making [5]. It is important to note that there is no single universal system of economic indicators that could be applied in all organizations. Besides, a lot of companies use varying assessment models of business based on analysis of various indicators. Design of a new intelligent interface, the ontological approach, was augmented by eye tracking methods. A visual presentation of knowledge and related data available in DSS allows for quick access to economic situation and for taking appropriate action [1]. The concepts and analytical operations obtained in study were matched with the financial ontology available in the system.

A balance sheet and profit and loss account of a real company were submitted for use in the research study. The financial statement included information indicating seemingly positive performance of the case company. However, in the periods analyzed, serious problems emerged as a result of debts. These problems should have been picked up by an analyst as threatening the loss of rationale behind going concern's basic assumption, i.e. the ability to function without threat of liquidation in the foreseeable future.

The experiment was divided into two phases. The first phase was to assess the overall situation of the company, based on core measures and indicators. After analyzing reports, each participant had to synthetically rate the state of company's assets and financial situation as very good, satisfactory, or poor. In the second phase, participants had to answer questions regarding financial indicators observed as positive and negative, with a higher level of detail. Detailed questions covered the key factors contributing to the development of the company, helping maintain its ability to continue its operations. Additionally, the purpose of this phase was to test ability to monitor adverse financial warnings which in many cases could have been concealed by the positive results and other financial indicators. Finally, participants had to comment on positive and negative aspects.

Data from experiments in the form of a log of operations along with ontology were analyzed in order to explore and produce an additional, often subconscious, knowledge on the manager. To analyze the data, algorithms of data mining and discovery sequential patterns were applied [7], however, it was necessary to prepare data beforehand.

III. Generate Sequence Using Eye Tracking

Two concepts are important in order to analyze eye movement, namely:
- fixations, that is, focus of eye vision on a particular item;
- saccades, which are sequences of rapid eye movements between successive fixations.

The points a manager focuses on can be identified based on the duration and number of fixations. In general, the shorter the time of eye fixation upon a specific item, the lesser significant the item is. The longer the time of fixation on an analyzed element, the more significant the element is in the process of perception [8].

Eye tracker and Gazepoint programming software were used for the experiment. They offer high-performance, affordable eye trackers to clients in all fields of applications from academic researchers to video game developers and more. This allowed for recording test footage, showing subsequent fixations and saccades for a defined research problem. Exemplary report in the eye tracker software of the case study is illustrated in Figure 2. This program offers us an opportunity to obtain different types of results on the visual activity of the participant. A heat map and an inverted heat map showing aggregate information on focusing of attention by the participant are some of these results that can be distinguished.

Numerous eye movements may be analyzed and processed to produce a variety of reports. Logs of the user operations provide detailed information on how information is perceived on the screen, on the order in which the content is viewed, and the length of time the eye is fixed on a given point. After transformation of the data, information on how a document was analyzed can be obtained.
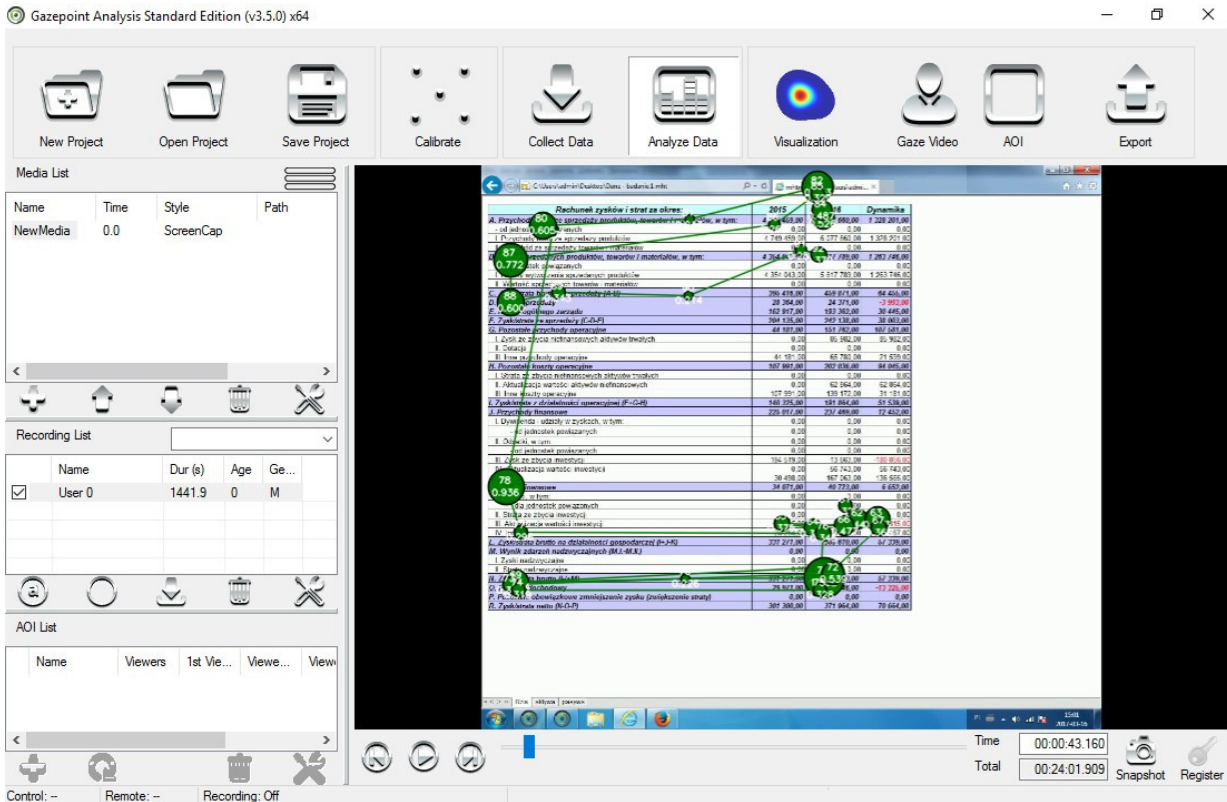
Fig. 2. Path of analysis financial statements

	1	2	3	4
101...	**Profit and loss account**	**2015**	**2016**	**Dynamics**
102...	A. Net revenues from sales of products, goods and materials, including:	4 749 459,00	6 077 660,00	1 328 201,00
103...	- from related parties	0,00	0,00	0,00
104...	I. net revenues from sales of products	4 749 459,00	6 077 660,00	1 328 201,00
105...	II. net revenues from sales of goods and materials	0,00	0,00	0,00
106...	B. Cost of products, goods and materials sold, including:	4 354 043,00	5 617 789,00	1 263 746,00
107...	- to related parties			
108...	I. Manufacturing cost of products sold	4 354 043,00	5 617 789,00	1 263 746,00
109...	II. Value of goods and materials sold	0,00	0,00	0,00
110...	C. Gross profit (loss) on sales (A-B)	395 416,00	459 871,00	64 455,00
111...	D. Selling costs	28 364,00	24 371,00	-3 993,00
112...	E. General and administrative costs	162 917,00	193 362,00	30 445,00
113...	F. Profit (loss) on sales (C-D-E)	204 135,00	242 138,00	38 003,00
114...	G. Other operating revenues	44 181,00	151 762,00	107 581,00
115...	I. Gain on disposal of non-financial fixed assets	0,00	85 982,00	85 982,00
116...	II. Subsidies	0,00	0,00	0,00
117...	III. Other operating revenues	44 181,00	65 780,00	21 599,00
118...	H. Other operating expenses	107 991,00	202 036,00	94 045,00
119...	I. Loss on disposal of non-financial assets	0,00	0,00	0,00
120...	II. Revaluation of non-financial fixed assets	0,00	62 864,00	62 864,00
121...	III. Other operating expenses	107 991,00	139 172,00	31 181,00
122...	I. Profit (loss) on operating activities (F+G-H)	140 325,00	191 864,00	51 539,00
123...	J. Financial revenues	225 017,00	237 469,00	12 452,00
124...	I. Dividend and profit sharing, including:	0,00	0,00	0,00
125...	- from related parties	0,00	0,00	0,00
126...	II. Interest, including:	0,00	0,00	0,00
127...	- from related parties	0,00	0,00	0,00
128...	III. Gain on disposal of investments	194 519,00	13 663,00	-180 856,00
129...	IV. Revaluation of investments	0,00	56 743,00	56 743,00
130...	V. Other	30 498,00	167 063,00	136 565,00
131...	K. Financial expenses	34 071,00	40 723,00	6 652,00
132...	I. Interest, including:	0,00	0,00	0,00
133...	- for related parties	0,00	0,00	0,00
134...	II. Loss on disposal of investments	0,00	0,00	0,00
135...	III. Revaluation of investments	5 815,00	0,00	-5 815,00
136...	IV. Other	28 256,00	40 723,00	12 467,00
137...	L. Profit (loss) on business activities (I+J-K)	331 271,00	388 610,00	57 339,00
138...	M. Result on extraordinary events (M.I. - M.II.)	0,00	0,00	0,00
139...	I. Extraordinary gains	0,00	0,00	0,00
140...	II. Extraordinary losses	0,00	0,00	0,00
141...	N. Gross profit (loss) (L+/-M)	331 271,00	388 610,00	57 339,00
142...	O. Income tax	29 971,00	16 646,00	-13 325,00
143...	P. Other statutory reductions in profit (increases in loss)	0,00	0,00	0,00
144...	R. Net profit (loss) (N-O-P)	301 300,00	371 964,00	70 664,00

Fig. 3. Split report into areas

Sequences, presentation of a scheme of analysis of financial statements played a key role in our study. As there was no easy way to acquire expected information from the system level, processing of data obtained was, arriving at specific sequences with accuracy of reports fields.

Details of data obtained are presented in Figure 3. Each report contains a given number of rows and columns. As a result of combination of these data, together with user scanning path, a sequence was obtained. The sequence provides information on method of analysis of SMEs reports.

The software provided by the eye tracker manufacturer allows for recording data on the log, with details on the recorded data. Figure 4 shows sample data that can be saved as a file. This has been described in same way in software documentation.

Among all the data, one of the most important in the attempt at defining areas of interest (AOI) is the information related to visual focus point, in relation to the X and Y axes. As these data are written on a scale of 0 to 1 showing the drawing, it is necessary to reference these values to resolution of the screen on which the study was conducted. In our case, it was 1680 x 1050 pixels for the X and Y axes, respectively. As a result, for example, if a participant looked closely at the center of the monitor, the data for both axes was 0.5, that is 840 and 525.

This information was used to define individual report fields as interest areas, where for each of them, on a proportional basis, a range of 0 to 1 was calculated for both axes. With the use of Excel and its functions IF, for each position, formulas were created to return information on whether a fixation described by log values is within defined AOI. Taking into account fixation number, duration and whether or not, in the area in which it occurred, a clear statement with a view of financial statement called the viewing sequence has been made.

Figure 5 shows, for example, generated viewing sequences for 5 different people. Numbers were assigned to relevant item in the financial statements, giving unequivocal information using numerical value:
- first - about the name of the report being analyzed;
- second and third - about the report line;
- fourth - about the report column.

Where there is 0 in the sequence, it means that a participant was not looking at a given fixation on the report, and his eyesight went beyond it. This may have been the result of either searching for data not on the report, such as system time or other software, or of fatigue, stress or dissociation. From the viewing sequence, it is hard to categorically pinpoint the cause, but testimony and recorded facial expression are helpful here. A detailed analysis of recordings will allow for more accurate identification of the reasons of the eyes straying outside the reports.

IV. SEQUENTIAL PATTERNS

Given a set of sequences where each sequence consists of a list of elements and each element consists of a set of items,

and given a user-specified min support threshold, sequential pattern mining consists of finding all of the frequent subsequences, i.e. subsequences whose occurrence frequency in the set of sequences is no less than min support [6].

Identification of sequential patterns of operations requires analysis of a database containing information about events that occurred within a specified period of time. Many methods used to detect sequential patterns have been applied in various fields, such as analysis of shopping cart, telecommunications, medicine, insurance banking, and Web pages [7].

Pattern recognition of analytical thinking may be performed using different algorithms. The choice depends on computational complexity as well as size of database. For the purposes of the study, only the most useful algorithms for detecting frequent sequential patterns will be examined.

The idea of an algorithm for discovering sequential patterns of is well-known [9]. First, minimum support for the occurrence of sequences in the database has to be defined. As a result, the algorithm produces in the first iteration frequent 1-element sequences for a given minimum support. In subsequent iterations, generating a candidate sequence uses the sequence discovered from a previous iteration. If the support of the discovered sequence is greater than or equal to predefined level of occurrences, the algorithm again generates a candidate sequence using the sequences detected in the previous iteration. Search algorithm ends when there are no more candidates or no common sequences.

Numerous modifications of algorithms to discover patterns of sequences resulted in elimination of identified difficulties and improved the efficiency of data mining. This is an example of data patterns conversion in the sequence shown in Figure 6. Conversion is a modification of the eye tracker logs arranged for fixation at a certain time. This made it possible to define occurrences at each AOI, which were then used to detect sequential patterns.

By comparing a manager's operations with the patterns of expertise and the domain ontology, the system is in a position to not only adapt the system interface to the level of the manager's knowledge, but also to identify "gaps" in his or her knowledge and propose an appropriate manner of economic analysis [10]. This will be possible using algorithms that identify dependencies and correlations in the way financial statements are analyzed. As a result, a detailed analysis of conclusions arrived at will allow for patterns generation and identify exceptions as well as their causes in the way reports are perceived. Pilot studies were also intended to provide information on opportunities to apply eye tracking tools to draw conclusions on the basis of sequential of perception of reports, and to define the rules and the process of realizing economic and financial expertise.
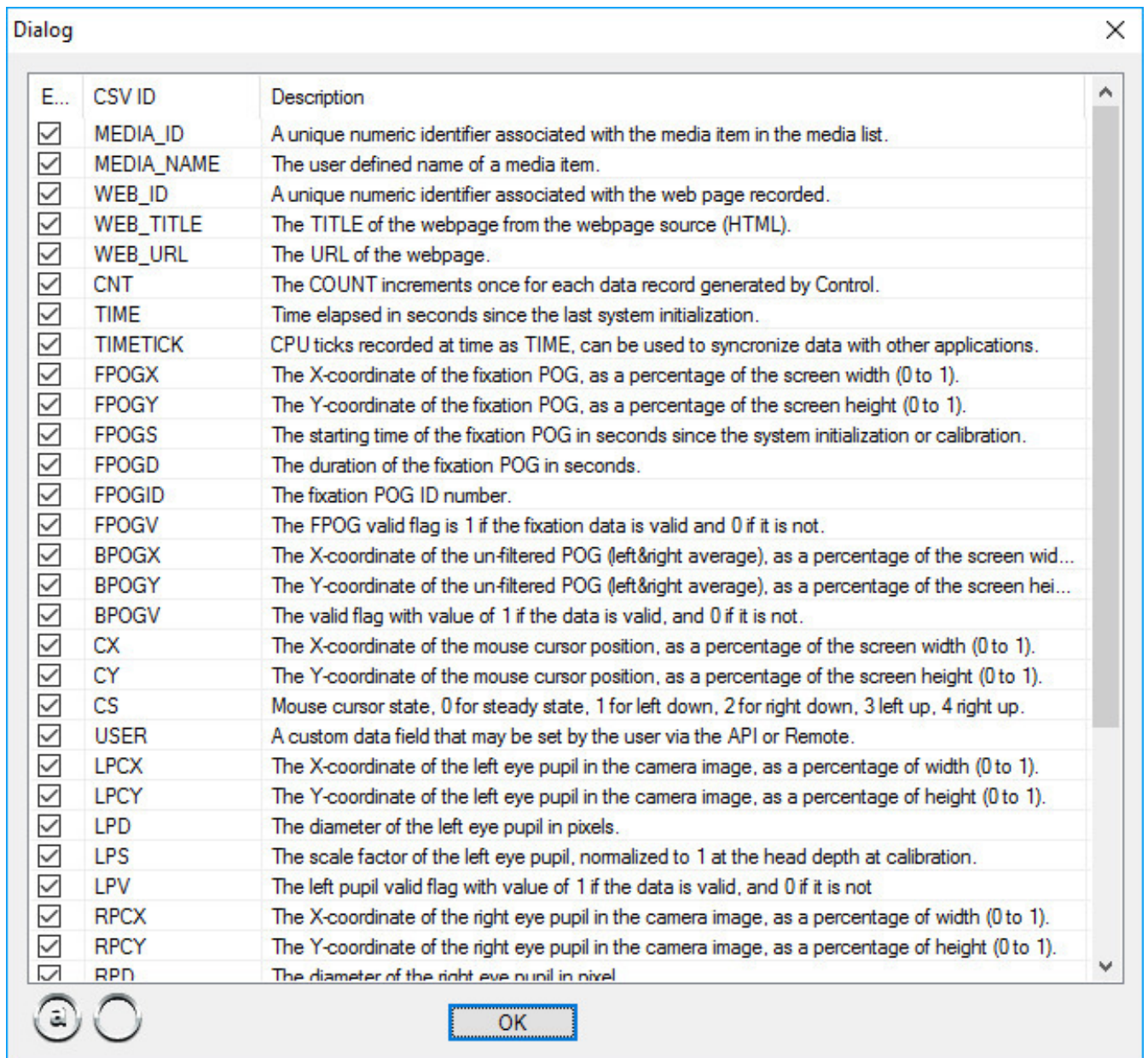
Dialog ✕

E...	CSV ID	Description
☑	MEDIA_ID	A unique numeric identifier associated with the media item in the media list.
☑	MEDIA_NAME	The user defined name of a media item.
☑	WEB_ID	A unique numeric identifier associated with the web page recorded.
☑	WEB_TITLE	The TITLE of the webpage from the webpage source (HTML).
☑	WEB_URL	The URL of the webpage.
☑	CNT	The COUNT increments once for each data record generated by Control.
☑	TIME	Time elapsed in seconds since the last system initialization.
☑	TIMETICK	CPU ticks recorded at time as TIME, can be used to syncronize data with other applications.
☑	FPOGX	The X-coordinate of the fixation POG, as a percentage of the screen width (0 to 1).
☑	FPOGY	The Y-coordinate of the fixation POG, as a percentage of the screen height (0 to 1).
☑	FPOGS	The starting time of the fixation POG in seconds since the system initialization or calibration.
☑	FPOGD	The duration of the fixation POG in seconds.
☑	FPOGID	The fixation POG ID number.
☑	FPOGV	The FPOG valid flag is 1 if the fixation data is valid and 0 if it is not.
☑	BPOGX	The X-coordinate of the un-filtered POG (left&right average), as a percentage of the screen wid...
☑	BPOGY	The Y-coordinate of the un-filtered POG (left&right average), as a percentage of the screen hei...
☑	BPOGV	The valid flag with value of 1 if the data is valid, and 0 if it is not.
☑	CX	The X-coordinate of the mouse cursor position, as a percentage of the screen width (0 to 1).
☑	CY	The Y-coordinate of the mouse cursor position, as a percentage of the screen height (0 to 1).
☑	CS	Mouse cursor state, 0 for steady state, 1 for left down, 2 for right down, 3 left up, 4 right up.
☑	USER	A custom data field that may be set by the user via the API or Remote.
☑	LPCX	The X-coordinate of the left eye pupil in the camera image, as a percentage of width (0 to 1).
☑	LPCY	The Y-coordinate of the left eye pupil in the camera image, as a percentage of height (0 to 1).
☑	LPD	The diameter of the left eye pupil in pixels.
☑	LPS	The scale factor of the left eye pupil, normalized to 1 at the head depth at calibration.
☑	LPV	The left pupil valid flag with value of 1 if the data is valid, and 0 if it is not
☑	RPCX	The X-coordinate of the right eye pupil in the camera image, as a percentage of width (0 to 1).
☑	RPCY	The Y-coordinate of the right eye pupil in the camera image, as a percentage of height (0 to 1).
☑	RPD	The diameter of the right eye pupil in pixel

ⓐ ◯   OK

Fig. 4. Data from recordings

Person	Sequence																			
1	0	1352	1291	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1011	
2	0	0	0	1442	1421	1393	1394	1402	1411	1421	1414	0	0	0	1031	1041	1091	1141	1161	
3	0	1321	1121	0	1012	1022	0	1091	1361	1041	1061	1081	1031	1041	1021	0	0	0	0	
4	1292	1301	1341	1411	1361	1381	1371	1342	1353	1354	1343	1351	1361	1321	1311	1291	1302	1431	0	
5	0	1331	1051	1031	1013	1044	0	1041	1081	1171	1151	1081	1051	1053	1044	1052	1012	1043	1084	

Fig. 5 Sequences of analysis financial statements

ParticipantName	LocalTimeStamp	EyeTrackerTimestamp	GazeEventType	GazeEventDuration	FixationPointX (MCSpx)	FixationPointY (MCSpx)
P02	09:49:10.627	1427787312203270	Fixation	267	143	237
P02	09:49:10.644	1427787312219940	Fixation	267	143	237
P02	09:49:10.661	1427787312236610	Fixation	267	143	237
P02	09:49:10.677	1427787312253270	Fixation	267	143	237
P02	09:49:10.694	1427787312269930	Fixation	267	143	237

ParticipantName	LocalTimeStamp	Number of fixations	Fixation1	Fixation2	Fixation3	Fixation4	Fixation5	Fixation6	Fixation7	Fixation8
P02	09:49:10.627	5	R	R	R	R	R			
P02	09:49:13.798	1	R							
P02	09:49:13.988	8	P	P	R	R	R	P	P	R
P02	09:49:16.690	3	R	R	R					
P02	09:49:18.190	4	P	P	P	P				

Fig. 6. Converting a point on the screen

V. CONCLUSION

In this project, a new part, referred to as Patterns of Expertise, is created on the basis of data provided by an eye tracker that records a manager's perception and his manner of document analysis. The collected quantitative data are used to identify sequential patterns of his perception of information. A built-in algorithm explores the eye tracker logs and creates a model of managerial operational knowledge.

Arriving at these data is make possible with the help of appropriately prepared data. As the eye tracking software doesn't have advanced data analysis functions, it was necessary to obtain a self-tracking tool. Interpretation of the data generated by the software, due the brief documentation, also constituted another problem. As a result, it was necessary to separately prepare the data. Their manual processing, as described, yielded the expected results. As a result, a database containing sequences was prepared. This will be used to apply pattern search algorithms, which will eventually be used to build intelligent system interfaces.

ACKNOWLEDGMENT

REFERENCES

[1] Korczak J., Dudycz H., Nita B., Oleksyk P., Kaźmieczak A., Attempt to extend knowledge of Decision Support Systems for small and medium-sized enterprises, Computer Science and Information Systems (FedCSIS), 2016, nr 8, pp. 1263-1271

[2] Korczak J., Kaźmierczak A., Discovery of Analytical Thinking Patterns of Managers using Eye Tracking, Intelligent Systems Conference (IntelliSys), 2017 [in print]

[3] Nita B., Korczak J., Dudycz H., Oleksyk P., Kaźmierczak A. (2015), W kierunku inteligentnych systemów analizy finansowej w zarządzaniu małym i średnim przedsiębiorstwem, Stowarzyszenie Księgowych w Polsce, Katowice, pp. 79-101

[4] Gruber T. (1995), Toward Principles for the Design of Ontologies Used for Knowledge Sharing, Academic Press, Duluth, pp. 907-928

[5] Dudycz, H.: The Topic Map as a Visual Representation of Economic Knowledge (in Polish). Wrocław University of Economics, Wrocław (2013)

[6] Agrawal R., Srikant R. (1995), Mining sequential patterns. Conf. Data Engineering (ICDE'95), Taipei, pp. 3–14

[7] Masseglia F., Teisseire M., Poncelet P. (2005), Sequential Pattern Mining: A Survey on Issues and Approaches, Institute of Electrical and Electronics Engineers, Gwalior, pp. 3-29

[8] Kamińska J. (2013), Okulografia w ocenie ergonomicznej stanowisk pracy – przykłady, Centralny Instytut Ochrony Pracy - Państwowy Instytut Badawczy. Warszawa, pp. 13-15

[9] Agrawal R., Srikant R. (1994), Fast Algorithms for Mining Association Rules, VLDB '94 Proceedings of the 20th International Conference on Very Large Data Bases, San Francisco, pp. 487-499

[10] Korczak J., Dudycz H., Nita B., Oleksyk P., Kaźmierczak A. (2015), Próba rozszerzenia wiedzy w systemach wspomagania decyzji menedżerskich w małych i średnich przedsiębiorstwach, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, pp. 227-241