

Bačun, Dinko; Lebarović, Nives Bačun

Conference Paper

A Model of Self-sustaining Automatic Corporate E-learning system

Provided in Cooperation with:

Governance Research and Development Centre (CIRU), Zagreb

Suggested Citation: Bačun, Dinko; Lebarović, Nives Bačun (2018) : A Model of Self-sustaining Automatic Corporate E-learning system, In: Tipurić, Darko Labaš, Davor (Ed.): 6th International OFEL Conference on Governance, Management and Entrepreneurship. New Business Models and Institutional Entrepreneurs: Leading Disruptive Change. April 13th - 14th, 2018, Dubrovnik, Croatia, Governance Research and Development Centre (CIRU), Zagreb, pp. 508-523

This Version is available at:

<http://hdl.handle.net/10419/180011>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

A Model of Self-sustaining Automatic Corporate E-learning system

Dinko Bačun, Nives Bačun Lebarović

Indicio d.o.o., Zagreb, Croatia

dinko@indicio.hr

nina@carpio.hr

Abstract

Challenging and volatile economic environment forced enterprises to seek alternative ways to optimize the cost and time needed to train their employees in adoption of new skills and technologies. Effective training shortens the time new skills are applied in practice, which results in shorter and more effective time to react to new market conditions. In the last two decades, technology advances revealed new possibilities. E-learning provided a large number of benefits over traditional classroom type training, providing consistent, world-wide training, increased learner convenience and lower expenses. This type of training proved to be slightly more effective than the classroom presentation. Its effectiveness grows with the number of participants. The savings that corporations reached, run in order of tens of millions of US dollars. However, practitioners highlight a number of drawbacks that actual implementations revealed. The upfront costs require considerable investment in both information technology and staff, as well as the hardware and software needed to run the system. The time needed to design and build the actual courses and professional knowledge necessary, implies lengthy implementation with sluggish response to change.

One aspect of corporate training is often overlooked and that is the training on internal procedures and best practices. When employee training is discussed in management circles, the courses to acquire new skills are considered in most cases. However, new employees need to be trained in corporate everyday practices, as well as experienced employees that are reassigned to new posts or to different departments. During the time they learn the processes, they are prone to errors which might go undetected for long periods of time and have serious consequences. Capturing the professional knowledge of everyday corporate practice proved to be a challenging task, as working professionals seldom have time to consistently describe the variations on prescribed corporate procedures. Traditionally, a tutor is assigned to introduce the newcomer to corporate everyday practices, but there is no measureable test that will reveal the level of knowledge adoption.

This paper describes a model that seamlessly captures everyday practices, analyses the distribution of their attributes and uses this information to create a set of questions/answers that are automatically fed into an e-learning platform with certification possibilities. Data is captured in an unrelated application in which the employees record everyday business processes. Captured data is analyzed periodically (weekly, monthly) by Analyzer module that produces distributions of occurrence of different process attributes, their steps and resources. Generation module applies rules to thresholds in the distribution to detect segments that are basis to generated questions/answers pairs which are fed to the certification platform. Changes in practice are detected as variations in distribution, which, when significant, can provoke creation of a new class with corresponding new questions/answers pairs rendering automatic modification of e-learning content.

Keywords: *Business process analysis, Corporate knowledge management, E-learning, Quiz generation*

Track: *Education*

Word count: 6.794

1. Introduction

For the past four decades enterprises captured an overwhelming amount of data with their computer systems. At first, used data was kept for record keeping purposes, but eventually it became clear that the data contains vast amount of knowledge that wasn't used in proactive way. As business processes became more automated, enterprise employees learned how to operate computers while the actual knowledge became more and more obscure, buried in sheer quantity of the data.

It soon became obvious that knowledge should be extracted from the data collected and this launched a worldwide search for data mining procedures. At first, the research focus was structured data, as it allowed for complex analysis of past situations which helped estimate possible future emerging issues.

After few years researchers became aware that large portion of data was kept in an unstructured form, as ordinary text, in documents and notes. A new approach was needed to understand the contents of such information. This led to numerous efforts to recognize natural language and take some action based on its analysis: Manning and Schütze (1999), Gentner and Goldin-Meadow (2003), Tanenhaus *et al.* (1995), Chowdhury(2003), Tan (1999), Sebastiani (2002).

Two scenarios emerged: in the first, a question is posted in natural language that has to be interpreted and a number of related documents presented as an answer. This is the action taken by the search engines. In the second scenario, the text is analyzed to generate questions about the content as an aid in learning process, which is the focus of this paper. Most procedures in this scenario rely on building parsing trees and apply different Natural Language Processing (NLP) procedures to get desired results. However, in everyday business processes, textual data captured is only a subset of full language, terse and to the point, but still free text style, unstructured data. No essays are written in corporate production. This allowed us to apply somewhat different approach and generate an E-learning system based on contents incidence of textual data.

E-learning research flourished in the last decade. Graf and List (2005) address issues of open source platforms, Ozkan and Koseler (2009) propose a conceptual assessment model, Khan (2005) the design and implementation of E-learning systems, Ardito *et al.* take first steps towards the definition of a methodology for evaluating e-learning applications, Strother (2002) discusses E-learning cost effectiveness in the corporate environment and Roffe (2002) the evaluation, quality assurance and engagement of operating an e-learning system.

Chali and Hasan (2015) conducted a research on topic to subtopic analysis and question generation. They assumed each topic is associated with a body of text having useful information about the topic. For the experiment they used the prepared set of 60 paragraphs

and about 60 topics. They generated 2,186 questions and evaluated top 15% and top 30% over four different systems. Three judges conducted the acceptability test, and the top 15% questions scored between 35.2% and 46.5%, while the top 30% performed lower.

Litman (2016) gives two example of NLP use in education technology: to automate the scoring of student texts with respect to linguistic dimensions such as grammatical correctness or organizational structure and to use dialogue technologies to achieve the benefits of human one-on-one tutoring.

Mitkov and Ha (2003) describe a computer-aided procedure for generating multiple-choice tests from electronic instructional documents. They found that 57% of the questions were judged worthy of keeping as test items.

Zhang and VanLehn (2016) compared human written questions to machine generated collected from the web and found out that that about 50 % of the human questions were deep whereas only about 30 % of the machine questions were deep.

Rus *et al.* (2007) researched the quality of questions generated by schemas from plain text. They found out that only 55% of generated question were rated acceptable.

Walsh *et al.* (2003) state that E-learning reduced corporate costs of training by tens of millions of dollars, but upfront costs are significant in information technology and staff as well as the cost of development of training materials.

Mandinach (2005) studied evaluation methodologies that address how E-learning is affecting teaching and learning process.

In the next sections, a model of an automated corporate E-learning system is described.

The methodology section is divided into subsections corresponding to different segments of the system: section 2.1 describes the analyzer procedures used to parse the input stream and prepare the input into the quiz generation process. Section 2.2 describes the quiz generator used to generate question/answer pairs. The exclusion criteria applied in the model used is discussed in section 2.3, while section 2.4 discusses the resulting framework. Section 2.5 describes how we tested the claim of self-sustainability and finally, section 2.6 describes the survey we used to verify the significance of the model.

2. Methodology

The goal of this paper was to test the effectiveness and relevance of an automated quiz generation system based on textual data collected by an unrelated corporate application and used to verify the knowledge of business processes in production. The input data used is raw, in the sense that it was not beautified or otherwise prepared before entering the knowledge extraction procedures, so no extra personnel effort was needed to maintain the E-learning system.

As a platform to test the hypothesis, we selected the actual data from the Internal Repair Workshop Unit in a large municipal road maintenance company, which is collected by an unrelated computer application, a Maintenance Repair Journal.

This project was done as part of a much larger Preventive Maintenance and Spare Parts Inventory Estimation. The same Analyzer procedures are used in both projects. The automatic quiz generation is used to assess the significance and level of accuracy that machine generated question / answer pairs can achieve. The Semantic Quiz Generator can fine tune the result sets to increase the relevance of the quiz data. Data analysis is performed periodically (daily, weekly, monthly) so that new developments automatically come into the

scope of the quiz data if their incidence is high enough, establishing a self-sustained certification platform.

When extracting knowledge from the machinery / vehicle repair process, there are two distinct situations. In the case of a single manufacturer, the data is coded structurally, the majority of it is well described in each model maintenance handbook and the relations, materials, hours and consecutive work necessary is declared. In this case, one would generate quizzes based on a fixed matrix of actions, so it would be easy to randomly pick a question and test a random number of correct answers.

In the case of a large corporate Internal Repair Workshop Unit, which repairs 977 different machines / vehicles, manufactured by more than twenty three different manufacturers and more than eighty two models, this is somewhat different. It is practically impossible and unnecessary to maintain a repair database for each manufacturer and model. Searching and extracting the proper process designation from any of multiple databases would certainly burden everyday repair workflow. In such cases, the repairs done are recorded as text, so that it would be easy for the employee to describe the repair done. Important particularities would automatically become part of the Maintenance History Log for each machine / vehicle.

The use of text in repair record keeping means that the knowledge to be extracted is based on textual data, which implies that lexical analysis needs to be done. Attempts to extract relevant meaning from lexical tokens has proved to be quite a challenge, and numerous authors have used neural network algorithms to achieve semantic correctness with feeble success. Fortunately, production data is not just free style text, but the records contain hidden relation meaning that enables selection of relevant sets of data, based upon set constraints.

When a defect is detected, the vehicle or machine is brought to the workshop, a Workshop Order is issued and all repairs recorded. Each Workshop Order records the machine / vehicle designation and origin which enables all the related data to be included into the analysis, like manufacturer, model, type, etc. A repair may consist of multiple jobs, described as 50 character textual Job Name field, generally categorized by type of work done (mechanical, electrical, etc.) which is used in cost analysis. The details of the job are described in 2500 character description field, where each work done is recorded as multiple sentences, referred to as Detailed Descriptions. When there was no Detailed Description field (like the WASHING job), the Detailed Description field was substituted by Job Name so that the relation between the two could be maintained.

The data considered for the purpose of this project consists of Workshop Orders No 846 to 2991 from the year 2017. The recording described above started on April 7th with the Workshop Order No 846 until Workshop Order 2991 issued on December 22nd.

The records were maintained on a daily basis by Internal Repair Workshop Unit staff, namely two foremen and the Unit manager. Each one of them has their own style of describing the work done, so in the input data may contain the same work described with different wording. Some of them used special characters (asterisks, exclamation points, etc.) to highlight a particular concept.

The analysis of the data was performed on corporate server, by the Advanced Lexical Analyzer module in January 2018. Live data was used as input. The Analyzer produced a set of relation tables described in the following section, which were fed as input to the Semantic Quiz Generator module. The Semantic Quiz Generator module generated question / answers pairs that were presented to the reviewers for rating.

2.1 Advanced Lexical Analyzer

The purpose of a lexical analyzer is to extract tokens (words) from a text. In our sample there were two such fields: the Job Name, a fifty character field, and the Detailed Description, a 2500 character field, organized as multiple sentences of work performed during a single repair job. The Advanced Lexical Analyzer performed a number of other tasks which results in a grid of relation data that enables more relevant quiz generation.

Beside those two fields, the input record contains other 261 fields (machine name, manufacturer, type of motor, etc.) which might be used to target analysis class.

For each Job Name, the analyzer calculates the total incidence of the particular job for the sample, as well as quarterly and monthly incidence, so that frequent period of the occurrence in the year can be evaluated.

For each Work Order the sequence incidence of two consecutive and three consecutive Job Names was calculated, total, quarterly and monthly. There are multiple jobs that are performed always consecutively and it is evident that their relevance in the total of jobs is significant.

Each Job Name was lexically broken into words, and the incidence of each word was calculated, as well as total, quarterly and monthly incidences. This would indicate significant verbs like repair, or replacement (as in tire replacement) to appear more often and add to the relevance of quiz generation.

For each word in the Job Name field, the consecutive order of two and three words incidence is calculated, total, quarterly and monthly. This would render focused incidence for relevant jobs.

Each Detailed Description was broken into sentences (Repair Actions) and for each Repair Action a total, quarterly and monthly incidence was calculated. Each Repair Action was broken into words and for each word the total, quarterly and monthly incidence was established. The consecutive order of two and three Repair Action words was calculated as total, quarterly and monthly.

The Analyzer further maintains relations between Job Names and Repair Actions, Job Names and Work Orders, Repair Actions and Work Orders, Job Name Words and Repair Actions and Repair Action Words and Repair Actions.

All the above incidences and relation analysis were performed on the total sample, which represents a single Class of Analysis we named Totals. However, it would be interesting if we could have such an analysis performed for each manufacturer, so that we could evaluate how one manufacturer compares for a particular repair to another. This resulted in twenty four sets of analysis by each of the manufacturers, and an extra analysis called Unknown for the machines with no manufacturer info entered into the database. This set of analysis formed a Class we called By Manufacturer and makes it possible to generate a question of the type: Which manufacturer has more X type repairs, A or B? The Class contains segments, in this case manufacturers, we call Seeds. The lexical analysis described above was performed for each Seed, for each manufacturer.

The Advanced Lexical Analyzer was developed as a tool, where user can analyze the selected data set by any field included in the input data stream. We analyzed a total of nine Classes:

- 1) Totals
- 2) By Manufacturer (26 Seeds)
- 3) By Each Vehicle (426 Seeds)

- 4) By Work Type (6 Seeds)
- 5) By Scheduled and Unscheduled Maintenance (2 Seeds)
- 6) By Vehicle Type (40 Seeds)
- 7) By Manufacturer Model (80 Seeds)
- 8) By Motor Type (33 Seeds)
- 9) By Business Units (16 Seeds).

2.2 Semantic Quiz generator

The Semantic Quiz Generator uses the data produced by the Advanced Lexical Analyzer to generate a set of question / answer pairs based on Question Schema. Each Schema may have an arbitrary number of Question Templates attached to it, which are used to create the actual text of the question. A Template contains ordinary text and keywords which are replaced by actual Token (2-token, 3-token combination), Job Name or Repair Action from the selected record.

There may be an unlimited number of Schemas. When a group of question / answer pairs is generated in a Batch, user specifies the total number of questions to generate, the Schemas that will be used in the generation and the percentage of the questions each Schema will generate. The Generator will then traverse each Schema and each Template in the Schema, pick a random question candidate and generate a question. On second pass, the Generator will pick another record with a different set of answer candidates to create the question. It will repeat this process until it reaches the Schema quota.

The Schema defines how the Generator should select the set of data from which it will pick a record to create a question. Once the question is selected, the answer data set can be only the one that is related to that particular question record, which eliminates the irrelevant answers. The relation tables give a set of plausible correct answers from which a predefined number of correct answers is randomly picked. Another instance of the same question will have a different set of correct answers. The incorrect answers, distractors, if they are to be included in the question response set, are picked from a pool of all other records.

A schema contains three segments: Questions, Answers and Templates. The question candidate set is defined in Schema Questions segment. It contains the Class and Seed from which the question will be picked. Then the source is defined: Job Names, Job Name tokens, Repair Actions, Repair Action tokens, 2-token or 3-token sequence of Job Name tokens, 2-token or 3-token sequence of Repair Action Words and finally Seed data. The user can further define the Lowest Incidence and Highest Incidence that the result set has to satisfy. The exclusion field is used to specify tokens that will force the record to be excluded from the question candidate set.

The Generator then randomly picks a record to create the question and replaces the keywords. It then takes the Schema Answer segment to determine the Correct Answer data set. The Answer segment contains the source Class and Seed which might be different from the question data set, allowing for cross-class responses. It also has the source selection with same options as the Questions folder, the Lowest and Highest incidence and non-inclusion token set.

The correct answer(s) can be picked in one of two ways: randomly or by incidence count. This is regulated the Use Incidence switch and adds variety to the system. The sorting order

(Ascending/Descending/None) is governed by Incidence Sort field. This allows the Generator to post questions of type: What are the two most common works done in exhaust repair job? The Template segment contains any number of templates which will generate the actual text of the question, so that a single Schema may generate different question texts. Each template definition contains text, combined with any number of keywords from the 263 fields of the input record or any of the relation tables. After the selected question candidate is paired to a correct answer candidate, all the keywords are known and can be easily replaced in the question text. The template definition also defines the number of correct answers and the number of distractors that each question should present to the user, as well as the number of questions each generation pass should generate. At last, it has a question type definition which can be: Multi-choice, Single-choice, Negative multi-choice, Negative single-choice and Yes/No.

2.3 Exclusion criteria

Three most common tokens (*and, in, for*) were skipped during token extraction so that their incidence does not influence token relations.

Both the schema Questions and Answers segments define the highest and lowest incidence that the question candidate set and its corresponding answer candidates have to satisfy. This was used to clear the candidate set of one time repairs which are not relevant for the knowledge of workshop processes. We used the same criteria to exclude the jobs that had incidence of one or up to 10, so that the question candidate set be reduced to about 20 to 30 candidates.

Both schema segments contain a field where exclusion tokens are defined. When a question candidate is picked, lexical analysis is performed and its words are tested against the exclusion tokens. If a match is found the question is skipped and another candidate is selected. We used this option to exclude the washing job, as it is part of almost any Workshop Order and has little relevance to maintenance process.

After the correct answer candidate set is selected, each candidate is tested for similarity with the question and rejected if it is the same. Further, the answer candidate is lexically analyzed and searched for any token found in selected question or question exclusion tokens. If a match is found, the correct answer candidate is rejected. In this way, the answers that would appear trivially correct to the selected question were avoided. If all the candidates for correct answers are rejected, a special answer "None of the above" is added and marked as correct. If there are less correct answers than the template dictates, the difference is released to the distractor pool.

The exclusion procedure for correct answer candidates is also applied to incorrect candidate set, so that there are no similar distractors that could cause ambiguity.

We decided to exclude the Totals class from quiz generation because we felt that this class would render too many generally positive results. We also excluded the questions form By Each Vehicle Class, although this class might render the most valuable knowledge for this particular Workshop, but we would lose the generality of the survey. The same reasoning mandated the exclusion of By Manufacturer Model Class. We also excluded By Work Type Class, because any electrical or mechanical repairs would be easily recognizable in the correct answers and those two types of repair form the majority of works done.

From the Classes chosen to be tested we discarded the seeds with low incidence so that we could generate multiple question/answer pairs from a single schema.

For simplicity sake, we used only Multi-choice type answer visual representation, so that the reviewers would easily recognize ambiguous answers.

We used the simplest of schema relations, namely Jobs-to-Works-done, declining to use 2-token, 3-token sources or cross-class sources although it would give much more precision.

2.4 Resulting generation framework

The resulting generation framework consisted of 20 schemas with one template each, generating questions from five different Classes and twenty seeds. Most of the questions generated 14 answers. We reasoned that the maximum we could expect the reviewers to evaluate, was two hundred question/answer pairs, which gave a pool of eight hundred questions. We wanted that each question/answer pair be evaluated by at least three judges, which meant that a total of 268 unique question must be generated. The second pass of the generation process distributed each question/answer pair to at least three different reviewers using a round-robin algorithm. The resulting questions were shuffled and distributed to forty quizzes with twenty questions each. Each reviewer was assigned consecutively ten different quizzes to evaluate.

2.5 Self-sustainability test

To test the system ability to recognize emerging new procedures, we introduced instances of a new job into the input stream, run it through Advanced Lexical Analyzer and generated a unique 268 question/answer pairs multiple times. For the test purposes we invented a job “Tachograph replacement” with two different works done in each instance. This job is actually done outside the Workshop, so there were no preexistent instances in the database. We opened a new Workshop Order and selected an Iveco truck whose appearance in the input stream would affect only three schemas, each with a different Low Incidence limit. We then analyzed the data set, run the generation procedure ten times and recorded the appearance of the new job in the resulting unique question/answer pairs. We repeated the same process for five, nine and fifteen new Work Orders with different trucks but otherwise same characteristics.

2.6 Survey

The model was tested in the working environment of the Internal Repair Workshop unit during January 2018. The Semantic Quiz Generator was used to generate 268 unique question / answers pairs which were fed into Survey data so that each question was rated by at least three reviewers. The questions and answer positions in each test were shuffled. The referee was presented with a machine generated question and machine selected answers, where some of them were marked as correct. He had to evaluate the meaningfulness of the question and whether the answers marked as correct were actually correct. The distractors presented would be the basis to evaluate ambiguity.

Evaluate the question and displayed answers

Evaluate the question and displayed answers

Evaluate this question

Nr: UID:

Question: U SERVIS spadaju radnje...

Select all the correct answers

Nr.	Answer	Correct?
1	IZMJENA NOSAČA MOTORA I KONVEKTORA	<input type="checkbox"/>
2	IZMJENA PRED FILTERA I FILTERA ZRAKA	<input checked="" type="checkbox"/>
3	IZMJENA ULJA U MOTORU	<input checked="" type="checkbox"/>
4	IZMJENA KABLA AKUMULATORA	<input type="checkbox"/>

Is this question relevant to the knowledge of business processes in the workshop

Relevance: Not relevant Slightly relevant Relevant

Is this question ambiguous, in the sense that it is formulated understandably, that it is clear what kind of answer is expected

Ambiguity: Ambiguous Not ambiguous

Figure 1: Sample Q/A pair presented to reviewers for evaluation: the question is: „A SERVICE includes...“ and the answers are: 1) Replacement of motor support and convector (not correct), 2) Replacement of forward filter and air filter (correct), 3) Replacement of oil in the motor (correct), 4) Replacement of battery cable (not correct).

We gave three choices for relevance: relevant, slightly relevant and not relevant (Figure 1). Most of the correct answer sets consisted of two correct answers, so if only one answer marked correct was actually correct, the question would be marked as slightly relevant. If both answers were correct but they were not strongly related to the question, it would be judged as not relevant. There were only two choices for ambiguity: ambiguous and not ambiguous.

Four knowledgeable participants were selected to evaluate the significance of the questions: the Internal Workshop Manager, two Workshop Foremen and Equipment Officer..

3. Results

The input data consists of 2146 Work Orders, but only 2118 orders were finished at the date the sample was taken. There were 3992 Jobs executed, out of which 1260 were unique. There were 6675 Detailed Descriptions recorded out of which 2194 were unique. A total of 564 tokens, 1150 2-tokens and 701 3-tokens were found in Job Names and 1597 tokens, 5735 2-tokens and 8104 3-tokens in Detailed Description.

In the review process, each referee evaluated a total of 200 question/answer pairs. The relevance could be assigned two (relevant), one (slightly relevant) and zero (non-relevant) points. The ambiguity could be assigned one (non-ambiguous) or zero (ambiguous) points. The average score for the entire set of 800 generated question/answer pairs is 72,38% for relevance and 83,50% for ambiguity. The scores for each referee is shown in Table 1.

	Relevance			Ambiguity		
	Max	Assigned	%	Max	Assigned	%
Referee 1	400	302	75.50%	200	198	99.00%
Referee 2	400	256	64.00%	200	151	75.50%
Referee 3	400	307	76.75%	200	189	94.50%
Referee 4	400	293	73.25%	200	130	65.00%
Totals	1600	1158	72.38%	800	668	83.50%

The performance by all the schemas is shown in Table 2.

The question/answer pair set was generated by twenty schemas. Column D shows how many unique question/answer pairs were generated from a particular schema. Initially, we started with 14 questions per schema and reduced the number to reach the 268 questions we needed for the quizzes. We reduced the Mercedes seed to ten, because it is the same manufacturer as MB seed, as the actual entry into the database was not consistent. We reduced the JCB and Bobcat seeds, because their average incidence was lower than those in other seeds. Column E shows how many question candidates were possible based on the schema question constraints. We reduced the constraints for some of the seeds which resulted with a larger number of possible questions, namely JCB, Mercedes, N2-Trucks and N3-Trucks because when Low Incidence was set to two, there were less than ten question candidates. When we removed this constraint, we got the shown number of candidates. Column F shows the possible number of answers to the first generated question. Each generated question has a different number of possible answers, so the number shown does not represent the number of all possible answers, but gives an indication of the quantity scale of answer candidates set. Column G shows schema relevance performance with all the judges. We summed the points assigned for relevance by the referees to all the questions generated by a particular schema and compared it to the maximum number of points a schema could score (column H). The

resulting percentage is shown in column I. The same calculation per schema was done for ambiguity schema score and those values are shown in columns J, K and L respectively.

Table 2

Class	Seed	Sch ema ID	Performance by Schema								
			Gener ated Q/A pairs	Possi ble quest ions	Poss ible ans wers	Relevance			Ambiguity		
						Poi nts	M ax	%	Poi nts	M ax	%
A	B	C	D	E	F	G	H	I	J	K	L
By Manufa cturer	MB	11	14	30	31	72	84	85. 71	40	42	95. 24
	IVECO	12	14	31	14	72	84	85. 71	37	42	88. 10
	JCB	13	12	62	31	54	68	79. 41	33	34	97. 06
	BOBCAT CATERPILL AR	14	10	27	8	39	60	65. 00	23	30	76. 67
		15	14	19	4	61	84	72. 62	33	42	78. 57
	MERCEDES	16	10	76	3	46	60	76. 67	24	30	80. 00
By Mainte nance Type	PLANNED	17	14	8	40	55	84	65. 48	36	42	85. 71
	UNPLANNED	18	14	34	28	18	84	21. 43	24	42	57. 14
By Vehicle Type	N3- TRUCKS	19	12	96	23	60	72	83. 33	33	36	91. 67
	N2- TRUCKS	20	14	73	8	58	84	69. 05	33	42	78. 57
	COMBO	21	14	21	48	58	84	69. 05	37	42	88. 10
	GRADER	22	14	31	40	62	84	73. 81	32	42	76. 19
	LOADER	23	14	27	9	56	84	66. 67	34	42	80. 95
By Motor Type	DIESEL	24	14	20	56	63	84	75. 00	36	42	85. 71
	DIESEL EURO II	25	14	11	14	67	82	81. 71	34	41	82. 93
	DIESEL- EURO III	26	14	10	12	67	82	81. 71	36	41	87. 80
	GASOLINE - EURO 2	27	14	16	8	67	84	79. 76	36	42	85. 71

By Business Unit	TRANSPORT	28	14	24	91	68	84	80. 95	35	42	83. 33
	MACHINERY WINTER MACHINERY	29	14	24	95	49	84	58. 33	34	42	80. 95
		30	14	10	13	66	84	78. 57	38	42	90. 48

Table 3 **Self-sustainability test**

Class	Seed	Schema ID	Low incidence limit	Incidence of a new job			
				1	5	9	15
A	B	C	D	E	F	G	H
By Manufacturer	IVECO	12	2	0	4	5	2
By Maintenance Type	UNPLANNED	18	10	0			3
By Business Unit	TRANSPORT	28	6	0		5	5

Table 3 shows the results of the test that was performed to understand whether appearance of new job procedures in the workshop business process would be followed by corresponding question/answer pairs in the resulting question/answer pairs. As stated in section 2.5 the new job was selected such, that only three different schemas 12, 18 and 28 could generate questions with the scope of the new job. Column D shows the low incidence limit of each schema. The job incidence has to be greater than this threshold to be included in the question candidate set. Column E shows the situation after the job was introduced for the first time in the input stream and ten question generation runs were invoked. Column F shows how many times was the question generated, after ten different question generation runs, following the new job introduction. Columns G and H show the results for nine and fifteen new Workshop orders containing the new job.

4. Discussion

Raw text data, from a working unrelated application, was fed into the system without any prior beautifying steps and analyzed. The result was a 72.38% relevant quiz system that can self-modify to react to new business procedures, creating a basis for a corporate E-learning system. The data was free hand text, keyed in by three different individuals, who were, at the time, not aware that their business process notes will be subject to analysis. We conducted interviews with the referees before the review process to make sure that they understood that they were evaluating the system performance, not their knowledge of the system or workshop business procedures.

Each employee had a different style in describing the work done and specifying the job name, so there were works recorded that were not strongly related to a particular job, like “transporting the tire to X business unit for vulcanization”. Although the notes were free style text, the vocabulary used was more limited than naturally speaking text, so it was reasonable to expect that an incidence based analysis may perform better than standard NLP algorithms.

This is understandable as no one writes large descriptions in production. Notes taken are rather terse and to the point.

We interviewed Referee No 2 to understand why his evaluation differs from the rest by 10%. He is a repair technology expert and he was expecting precise formal answers. For question about “the repair of electrical installation”, he would expect answers related to the repair of wiring, so that the answers “repair of signaling” and “replacement of the main cable” were evaluated as non-relevant, although they were specified as work done in the actual Workshop Order. In the case when all the correct answer candidates were discarded due to exclusion rules, the correct answer “None of the above” was voted irrelevant in 18 (9.00%) cases. There were 24 such questions (12.00%) in his questions set. Changing schema parameters would reduce this number. His duties do not include specification of works done and thus his judgement was conservative but he still voted 64% question/answer pairs as relevant to the knowledge of the workshop business process.

The schema 18, Unplanned Maintenance, had by far the lowest performance of only 21.43% of relevance. We researched the answer data set for the question candidates. It is a rather vast group of works, performed when a machine or a vehicle is brought into the workshop and all kind of works, that were long due, are done with a job name that is rather unspecific. So when such a question is picked, the answer set gives answer pairs that, although actually performed, are not strongly related to the question. This could be easily remedied by defining answer set Low Incidence constraint.

No optimizing steps during analysis and quiz generation were taken, as the main goal was to test the plausibility of an automated self-sustained corporate e-learning system. Classes that were too general were avoided, as we felt that they would produce quizzes with little relevance, as was observed in Unplanned Maintenance Seed.

However, the Unplanned Maintenance Seed performance could be much improved with the focus on a particular group of tokens. Schemas based on token selection were not used in quiz generation runs. The schema definitions allow the definition of question candidate set based on a particular token or a 2-token (or 3-token) combination. For example, there are 60 jobs starting with the word “Control”, but there are only 39 2-token combinations that are meaningful. It is easy to discard non meaningful combinations because the second token has much lower incidence. The same applies to the answer candidate set, as the Detailed Descriptions of works done is also tokenized. Even if a number of unimportant works were specified under a particular job name, the incidence of the answer tokens (2-token, 3-token) will reveal the dependencies that are meaningful.

There may be questions that deserve special focus. If a particular job is of special importance, like “Preparation for technical inspection” all the important works can be singled out by fine tuning the schema.

Multiple schemas can exist for each Seed in a Class. Fine tuning the schemas by a technology expert designer will render highly relevant quizzes.

A schema may address incidence data across classes allowing for a whole new variety of quiz questions. The quarterly and monthly incidences allow questions in the area of seasonal variations.

The multiple generation runs show that the system is capable of self-modification to address new developments without additional user intervention. The results in Table 3 show that the number of times that a new job appears in the input stream, has little influence on question incidence, as the questions are picked randomly. It is easy to imagine such a system

performing periodical analysis automatically, inquiring participants' knowledge of business processes.

The system described is independent of the input data. It takes whatever input stream is defined and applies the Advanced Lexical Analysis and Semantic Quiz Generator runs to it. It would be easy to define a new view of the input stream from a different source. If there are two textual fields that are in described relation, any other input field may be target of a Class. Careful selection of Class targets will render an automated corporate E-learning system with the same level of relevance.

5. Conclusion

This paper presents a model of self-modifying corporate E-learning system based on analysis of raw textual data obtained from an unrelated production application which records everyday business processes in an Internal Repair Workshop. Two textual fields from the input stream are broken into single or multiple tokens, their incidences are counted and later used to determine relevant question/answer pairs in the quiz generation step. All the other input fields became possible targets of a new Class of analysis. There may be unlimited number of Classes. Schemas and templates govern model behavior. Their number per Class is not limited.

Although the input stream consists of free style text, input by three different individuals, the business process data represents only a subset of natural language text, so that analysis and question/answers generation based on incidences rendered 72.38% relevant acceptance compared NLP algorithms (35% - 50%)

If the analysis step is configured as an automatic periodic computer task, the system becomes self-sustaining, recognizing newly introduced processes without any further intervention and generate questions that query participants' knowledge of new developments.

References

- Ardito C *et al.* 2006. An approach to usability evaluation of e-learning applications. *Universal access in the information society*, 4(3), 270-283.
- Chali, Y, Hasan, SA. 2015. Towards topic-to-question generation. *Computational Linguistics*, 41(1), 1-20. https://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00206
- Chowdhury GG. 2003. Natural language processing. *Annual review of information science and technology*, 37(1), 51-89.
- Sebastiani F. 2002. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1), 1-47.
- Gentner D, Goldin-Meadow S. 2003. Language in mind: Advances in the study of language and thought. *MIT Press*.

- Graf S, List B. 2005, July. An evaluation of open source e-learning platforms stressing adaptation issues. *In Advanced Learning Technologies, 2005. ICALT 2005. Fifth IEEE International Conference on* (pp. 163-165). IEEE.
- Khan BH. 2005. Managing e-learning: Design, delivery, implementation, and evaluation. *IGI Global*.
- Litman DJ. 2016. Natural Language Processing for Enhancing Teaching and Learning. *In AAAI* (pp. 4170-4176).
- Mandinach EB. 2005. The development of effective evaluation methods for e-learning: A concept paper and action plan. *Teachers College Record*, 107(8), 1814.
- Manning CD, Schütze H. 1999. Foundations of statistical natural language processing. *MIT press*.
- Mitkov R, Ha LA. 2003, May. Computer-aided generation of multiple-choice tests. *In Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing-Volume 2* (pp. 17-22). Association for Computational Linguistics.
- Ozkan S, Koseler R. 2009. Multi-dimensional students' evaluation of e-learning systems in the higher education context: An empirical investigation. *Computers & Education*, 53(4), 1285-1296.
- Roffe I. 2002. E-learning: engagement, enhancement and execution. *Quality Assurance in Education*, 10(1), 40-50.
- Rus V, Cai Z, Graesser AC. 2007, February. Experiments on generating questions about facts. *In International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 444-455). Springer, Berlin, Heidelberg.
- Strother JB. 2002. An assessment of the effectiveness of e-learning in corporate training programs. *The International Review of Research in Open and Distributed Learning*, 3(1).
- Tan AH. 1999, April. Text mining: The state of the art and the challenges. *In Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases* (Vol. 8, pp. 65-70). sn.
- Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.

Welsh ET, Wanberg CR, Brown KG, Simmering MJ. 2003. E-learning: emerging uses, empirical results and future directions. *International Journal of Training and Development*, 7(4), 245-258.

Zhang L, VanLehn K. 2016. How do machine-generated questions compare to human-generated questions?. *Research and Practice in Technology Enhanced Learning*, 11(1), 7.