Creating Contract Templates for Car Insurance Using Multi-Agent Based Text Understanding and Clustering

Igor Minakov [1], George Rzevski [2],
Petr Skobelev [1], Simon Volman [1]
[1] Magenta Development, 1a Osipenko St., Samara, 443010, Russia
{minakov, skobelev, volman}@magenta-technology.ru
[2] Magenta Technology, Gainsborough House, 59 - 60 Thames Street, Windsor, Berkshire
Sl4 1TX, UK
george.rzevski@magenta-technology.com

Abstract. The paper discusses the problem of automated processing and classification of unstructured text and proposes a new approach based on the multi-agent technology. The proposed method was successfully used to develop a system for a large UK insurance company capable of analysing and classifying 25000 documents related to car insurance, as well as creating template contract documents. The paper describes the system, presents testing results and discusses perspectives.

Keywords: template building, problem domain ontology, multi-agent text understanding, semantic network comparison, text mining, data mining, classification, clustering

1. Introduction and Problem Definition

Contemporary companies work with a very large number of documents, dealing with contracts, emails, business letters, licenses, manuals, financial and technical reports etc. Even for a medium-size company, the number of documents is so high, that it is impossible to process them manually without electronic document circulation or CRM systems. In addition to performing usual tasks, there is often a need for tools for deeper analysis, including semantic search and comparison of documents, grouping of documents with similar meanings and even an automatic document generation.

The method and supporting system described in this paper was developed for a client, one of the five biggest insurance companies in the UK, which had the following problem. Since car insurance premiums depend on many parameters, including client's gender and age, their education level, yearly income, class of the car and driving history, their lawyers created over the last 20 years more than 25,000 documents related to car insurance contracts. The task given to our team was to analyse these documents, classify them according to their semantic similarity and create a contract template for each group of documents. Contract templates were expected to include the most frequent clauses from the constituent documents within the group to be used in future as a basis for all new contracts. A part of the task was also to analyse and classify available contracts from the competitive insurance companies, and take these results into account during the creation of templates. The initial estimate was that there would be around 100 groups of documents, and that the whole work would take approximately 16 man-years of highly qualified law experts. Our task was to automate this process thus saving time and money.

2. The Solution

The existing systems and algorithms, which are capable of clustering documents (even the best ones, like LSA [1], Scatter/Gather [2] and STC [3]), have a number of limitations [4], as follows. Some of them require "seed" document in each group, some require expert pre-analysis, including desired number of resulting groups and others produce inadequate results because they use keywords-based rather than semantic searches and therefore produce a considerable noise and irrelevant results. To the best of our knowledge there are no other algorithmic or multi-agent based methods that could handle text understanding, clustering and template creation.

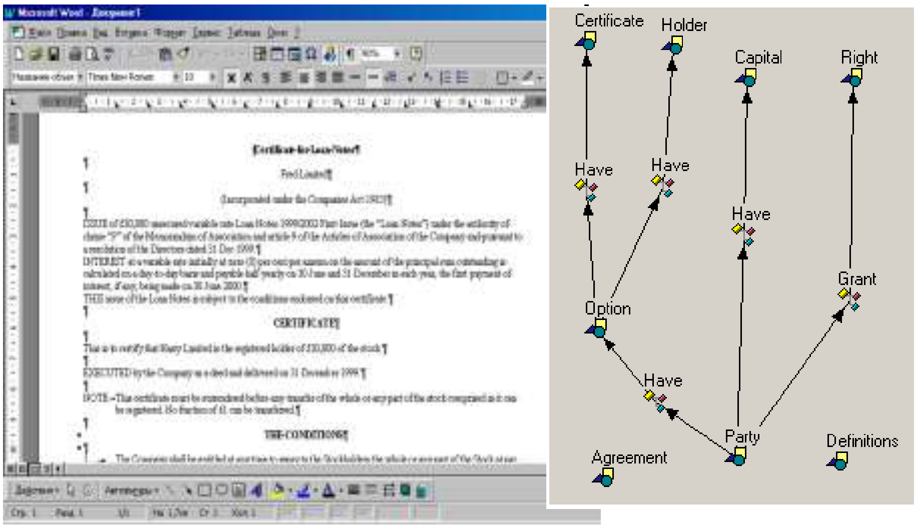
Therefore we have used Magenta multi-agent technology for all tasks of the requirement specification. First, our previously developed ontology and multi-agent based method for text understanding was used to represent the document meaning as a semantic network. Then, our multi-agent clustering method was applied to classify documents; and, finally, a heuristic method was developed to create templates based on groups of semantically similar documents.

2.1. Ontology

The proposed method is knowledge-based rather than data-driven. Conceptual knowledge of the domain of car insurance is contained in Ontology, which is constructed in a semi-automatic way, as follows. Using the ontology constructor [5] an expert inputs a set of domain documents (in our case – insurance contracts), and the system suggests hierarchy of objects, attributes and relations, which the expert can adjust manually. The heuristics are similar to those, used in [6] and [7]. For the car insurance domain ontology includes more than 400 objects (like "document", "agreement", "contractor", "terms of contract"), on the average, six attributes per object (like "amount", "class of car", "car parameters", "contract lengths", "warranty conditions" etc), and 37 relations (like "have", "between", "part of contract", "belongs to", "guarantee" etc).

2.2. Semantic Analysis

Using the domain knowledge from ontology, a multi-agent text understanding engine, described in [8, 9], assigns to each document a Semantic Descriptor, which describes the meaning of the document in the form of a semantic network (Fig. 1). The expression "meaning of the document" is used here to denote the meaning of the most relevant and important information contained in the document. For example, from the fragment of a semantic network shown in Fig. 1 it is possible to deduce that it describes a certificate for options, which belongs to a party of the agreement and that options belong to the registered holder. By similar analysis it is quite easy to interpret any semantic network and to determine the meaning of the document and, therefore, its possible future use. Also to be able to use standard clustering methods, in addition to semantic descriptors, keywords and key clauses are created for each document.



Certificate for Loan Notes

Fred Limited

(Incorporated under the Companies Act 1985)

ISSUE of £50,000 unsecured variable rate Loan Notes 1999/2002 First Issue (the "Loan Notes") under the authority of clause "5" of the Memorandum of Association and article 9 of the Articles of Association of the Company and pursuant to a resolution of the Directors dated 31 Dec 1999.

INTEREST at a variable rate initially at zero (0) per cent per annum on the amount of the principal sum outstanding is calculated on a day-to-day basis and payable half yearly on 30 June and 31 December in each year, the first payment of interest, if any, being made on 30 June 2000.

THIS issue of the Loan Notes is subject to the conditions endorsed on this certificate.

CERTIFICATE

This is to certify that Harry Limited is the registered holder of £10,000 of the stock

EXECUTED by the Company as a deed and delivered on 31 December 1999

NOTE – This certificate must be surrendered before any transfer of the whole or any part of the stock comprised in it can be registered. No fraction of £1 can be transferred.

THE CONDITIONS

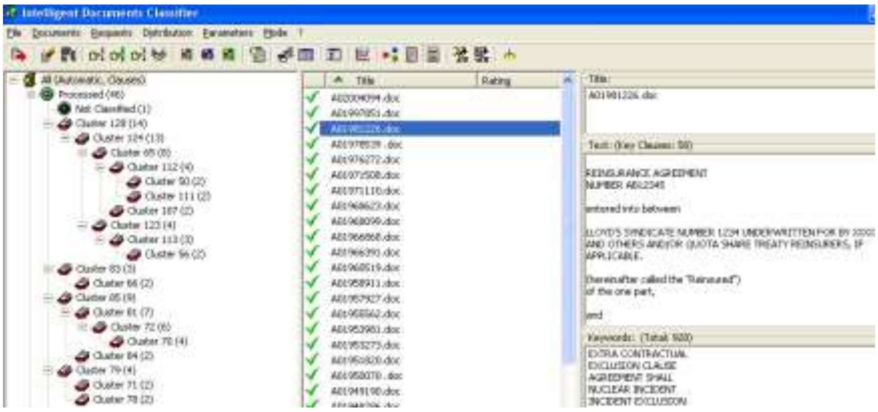
The Company shall be entitled at any time to [illegible] the Stockholders the whole or any part of the Stock at par

Fig. 1. A fragment of a semantic descriptor for an insurance contract

2.3. Clustering of Documents

Once semantic descriptors for all documents have been built, a multi-agent clustering engine, described in [10, 11], creates clusters of documents with similar meaning. The use of semantic descriptors helps to interpret reasons behind grouping documents together.

As a result we get a hierarchical structure where each document, or cluster of documents, can belong to several clusters, the criterion being the similarity of their semantic descriptors, as shown in (Fig. 2).



All (Automatic, Clauses)
Processed (46)
Not Classified (1)
Cluster 128 (14)
Cluster 124 (13)
Cluster 65 (8)
Cluster 112 (4)
Cluster 90 (2)
Cluster 111 (2)
Cluster 107 (2)
Cluster 123 (4)
Cluster 113 (3)
Cluster 96 (2)
Cluster 83 (3)
Cluster 66 (2)
Cluster 85 (9)
Cluster 81 (7)
Cluster 72 (6)
Cluster 70 (4)
Cluster 84 (2)
Cluster 79 (4)
Cluster 71 (2)
Cluster 78 (2)

✓ A02004094.doc
✓ A01997851.doc
✓ A01981226.doc
✓ A01978539.doc
✓ A01976272.doc
✓ A01971508.doc
✓ A01971110.doc
✓ A01968623.doc
✓ A01968099.doc
✓ A01966868.doc
✓ A01966393.doc
✓ A01960519.doc
✓ A01958911.doc
✓ A01957927.doc
✓ A01955562.doc
✓ A01953981.doc
✓ A01953273.doc
✓ A01950820.doc
✓ A01950070.doc
✓ A01949190.doc

Title:
A01981226.doc

Text: (Key Clauses: 50)

REINSURANCE AGREEMENT
NUMBER A012345

entered into between

LLOYD'S SYNDICATE NUMBER 1234 UNDERWRITTEN FOR BY XXXX
AND OTHERS AND/OR QUOTA SHARE TREATY REINSURERS, IF
APPLICABLE.

(hereinafter called the "Reinsured")
of the one part,

and

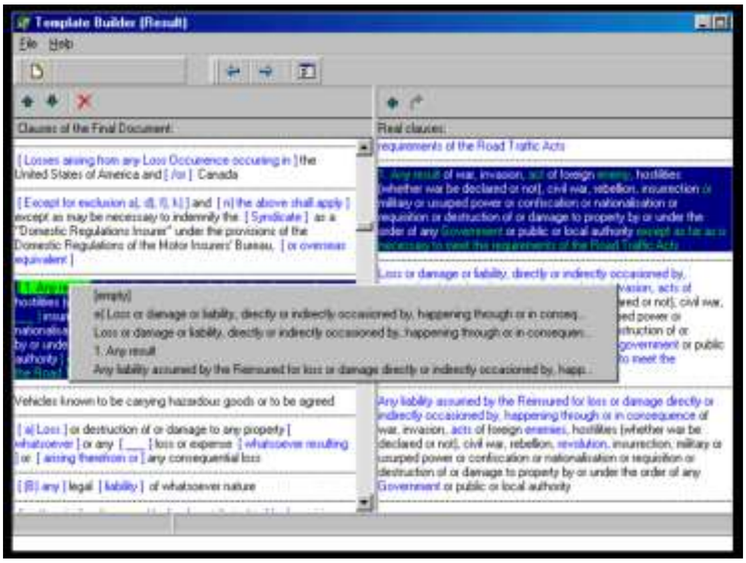
Keywords: (Total: 920)
EXTRA CONTRACTUAL
EXCLUSION CLAUSE
AGREEMENT SHALL
NUCLEAR INCIDENT
INCIDENT EXCLUSION

Fig 2. An example of document clusters for car insurance

The multi-agent clustering engine represents each cluster as a record with number of key clauses, which describe the cluster. A procedure of comparing clauses takes into account number of similar words in clauses and their corresponding order. Clauses with high degree of similarity are considered to be the same. The most frequent clauses among documents in a set are labelled as key clauses.

2.4. Creating Templates

Clusters produced by the clustering engine are analysed with a view to finding the most popular clauses, similar clauses and abnormalities for each group of clusters. All key clauses, which are popular and unique/abnormal are joined together to form the final template. To determine the order in which they should appear in the resulting document, a dynamic programming algorithm is used, which considers order in which clauses appear in all documents in the group (Fig. 3). Results are passed to experts who can re-adjust the order of clauses, select options or edit words in partially matched clauses (where similar clauses have some differences in wording) or include additional clauses.



[Losses arising from any Loss Occurence occuring in] the United States of America and [/or] Canada

[Except for exclusion a), d), f), k)] and [n) the above shall apply] except as may be necessary to indemnify the [Syndicate] as a "Domestic Regulations Insurer" under the provisions of the Domestic Regulations of the Motor Insurers' Bureau. [or overseas equivalent]

[empty]
e) [Loss or damage or liability, directly or indirectly occasioned by, happening through or in conseq...
Loss or damage or liability, directly or indirectly occasioned by, happening through or in consequen...
1. Any result
Any liability assumed by the Reinsured for loss or damage directly or indirectly occasioned by, happ...

Vehicles known to be carrying hazardous goods or to be agreed

[a) Loss] or destruction of or damage to any property [whatsoever] or any [___] loss or expense [whatsoever resulting] or [arising therefrom or] any consequential loss

[B) any] legal [liability] of whatsoever nature

requirements of the Road Traffic Acts

1. Any result of war, invasion, act of foreign enemy, hostilities (whether war be declared or not), civil war, rebellion, insurrection or military or usurped power or confiscation or nationalisation or requisition or destruction of or damage to property by or under the order of any Government or public or local authority except as far as is necessary to meet the requirements of the Road Traffic Acts

Loss or damage or liability, directly or indirectly occasioned by,
vasion, acts of
ared or not), civil war,
ed power or
struction of or
government or public
to meet the

Any liability assumed by the Reinsured for loss or damage directly or indirectly occasioned by, happening through or in consequence of war, invasion, acts of foreign enemies, hostilities (whether war be declared or not), civil war, rebellion, revolution, insurrection, military or usurped power or confiscation or nationalisation or requisition or destruction of or damage to property by or under the order of any Government or public or local authority

Fig. 3. Creating a template based on a set of key clauses from a cluster of documents

3. Multi-Agent Technology

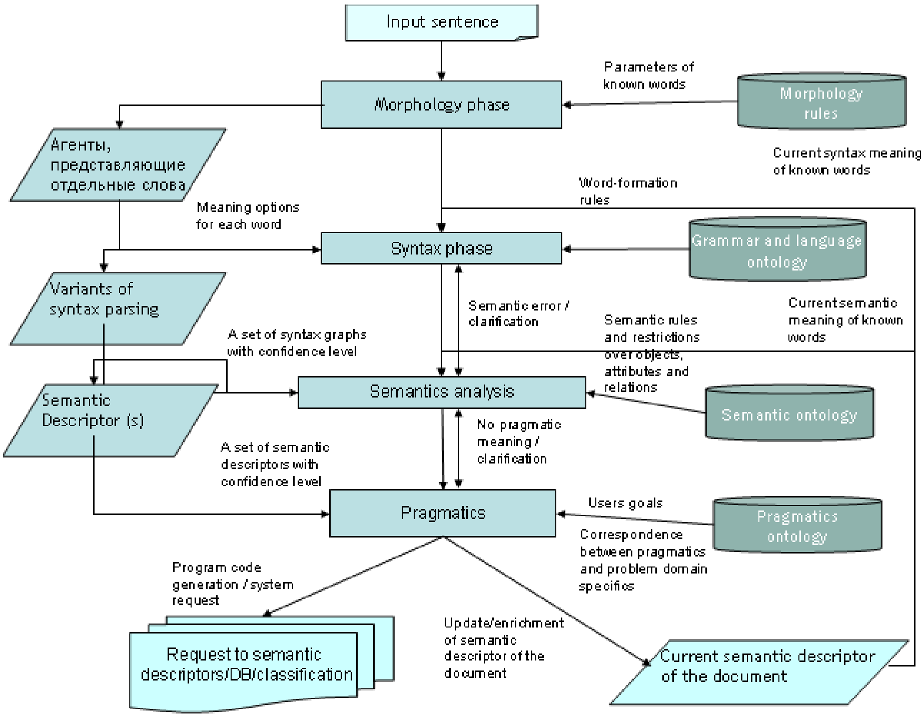
The multi-agent technology employed for semantic analysis (text understanding) and clustering is described here briefly to enable the reader to comprehend the logic of the method. A more detailed description can be found in [9, 10] and a detailed comparison with other methods, alongside with performance analysis and number of real-life applications can be found in [7, 11]. It is worth mentioning, that multi-agent applications are designed following holistic approach by building open Resource and Demand Networks, where demand agents and resource agents can join or leave the network dynamically, in real time [12]. In this application Demand and Resource

Agents are assigned to records and clusters for clustering, and to words and their possible meanings for text understanding, respectively. The input situation, called the initial Scene, is formally represented as a network of instances of concepts from ontology. The semantic analysis engine and clustering engine are event driven. Results of ongoing matching and re-matching of resources to demands are represented as intermediate scenes and the final result as the final scene.

3.1 Multi-agent Text Understanding

The method consists of the following four steps (see Figure 4): (1) Morphological Analysis; (2) Syntactic Analysis; (3) Semantic Analysis; (40 Pragmatics.

The process is as follows. The whole text is divided into sentences. Sentences are fed into the meaning extraction process one by one and the first three stages are applied to each sentence. After the text is parsed, the resulting semantic descriptor enters the forth stage – pragmatics.



Input sentence
Morphology phase
Parameters of known words
Morphology rules
Агенты, представляющие отдельные слова
Current syntax meaning of known words
Word-formation rules
Meaning options for each word
Syntax phase
Grammar and language ontology
Variants of syntax parsing
Semantic error / clarification
Semantic rules and restrictions over objects, attributes and relations
Current semantic meaning of known words
A set of syntax graphs with confidence level
Semantic Descriptor (s)
Semantics analysis
Semantic ontology
No pragmatic meaning / clarification
A set of semantic descriptors with confidence level
Pragmatics
Users goals
Pragmatics ontology
Correspondence between pragmatics and problem domain specifics
Program code generation / system request
Update/enrichment of semantic descriptor of the document
Request to semantic descriptors/DB/classification
Current semantic descriptor of the document

Fig. 4. General scheme of the multi-agent text understanding method

Morphological Analysis - An agent is assigned to each word in the sentence; (1) Word Agents access ontology and acquire relevant knowledge on morphology; (2) word agents execute morphological analysis of the sentence and establish characteristics of each word, such as gender, number, case, time, etc.; (3) if morphological analysis results in polysemy, i.e., a situation in which some words could play several roles in a sentence (a noun, an adjective or a verb), several agents are assigned to the same word each representing one of its possible roles thus creating several branches of possible sentence parsing results.

Syntactical Analysis – (1) Word agents access ontology and acquire relevant knowledge on syntax; (2) word agents execute syntactical analysis where they aim at identifying the syntactical structure of the sentence; for example, a subject searches for a predicate of the same gender and number, and a predicate looks for a suitable subject and objects. Conflicts are resolved through a process of negotiation; a grammatically correct sentence is represented by means of a syntactic descriptor; (3) if results of the syntactical analysis are ambiguous, i.e., several variants of the syntactic structure of the sentence under consideration are feasible, each feasible option is represented by a different syntactic descriptor; if no syntactic descriptor is valid, then several options with high enough level of correctness (i.e. more than 80% of grammatically possible links are present in the syntactic descriptor) are selected for the next stage.

Semantic Analysis - (1) Word agents access ontology and acquire relevant knowledge on semantics, including possible relations between concepts and valid attribute values; (2) each selected grammatical structure of the sentence under consideration is subjected to semantic analysis; this analysis is aimed at establishing the semantic compatibility of words in each grammatically correct sentence; from the ontology word agents learn possible meanings of words that they represent and by consulting each other attempt to eliminate inappropriate alternatives by building least contradictive semantic descriptor based on the problem domain ontology; (3) once agents agree on a grammatically and semantically correct sentence, they create a semantic descriptor of the sentence, which is a network of concepts and attribute values contained in the sentence; (4) if a solution that satisfies all agents cannot be found, agents compose a message to the user explaining the difficulties and suggesting how the issues could be resolved or, if the level of correctness is high enough (i.e. more than 80% links are valid according to the problem domain ontology), agent select most probable decision autonomously; (5) each new grammatically and semantically correct sentence generated in the previous steps is checked for semantic compatibility with semantic descriptors of preceding sentences; in the process agents may decide to modify previously agreed semantic interpretations of words or sentences by returning to earlier stages of negotiation (described above) with their new knowledge; this may improve confidence in certain options and may result in the reconstruction of semantic descriptors for preceding sentences; (6) when all sentences are processed, the final semantic descriptor of the whole document is constructed thus providing a computer readable semantic interpretation of the text.

3.2. Multi-Agent Clustering Method

The method is conceptually simple and elegant. An agent is assigned to every data element (record, document, text segment) and given the task to seek similar data elements with a view to forming a cluster. Agents fulfil their tasks by sending invitations to other agents and by responding to received invitations. Each agent is looking after its own interests and will join other agents in a cluster only if it suits its objectives. Once a cluster is formed an agent is assigned to it with the task to attract suitable data elements. Agents of data elements belonging to a cluster form temporary virtual communities, which can be organised in many different ways.

In a simple case, records try to find clusters with the maximum density. In more complex cases, metrics can include number of records, number of sub-dimensions for a cluster, time of life in the cluster, type of attributes etc. The search always starts with the nearest candidate and extends gradually. When a record finds a proper cluster, it makes an offer and waits for a reply. The cluster considers record's locality, calculates its variant and either accepts or rejects the offer. Thus instead of a centralised optimal top-down decision of a classic clustering algorithm, the multi-agent approach builds solutions in a stepwise manner, bottom-up. The matching of records to clusters is based on a current local balance of interests of a particular record and a particular cluster. If both parties agree, the record enters the cluster, if not – the record searches for other options.

The stages of dynamic clustering process for a simple 2D case are shown in Fig. 5. Stage a – the first record arrives. Stage b – the second record arrives; the two records form a cluster. Stage c – the third record arrives; the first cluster and the third record form another cluster. Stage d – the second cluster invites the records from the first cluster to join it; records decide to accept invitation because the switch offers them certain advantages; the first cluster is dismantled; the fourth record arrives. Stage e – a new cluster is formed. Stage f – a new cluster invites the records from the inner cluster to join it and they accept; a new record arrives. Stage g – a new record arrives and the process is repeated. Stage h – the final cluster is formed.

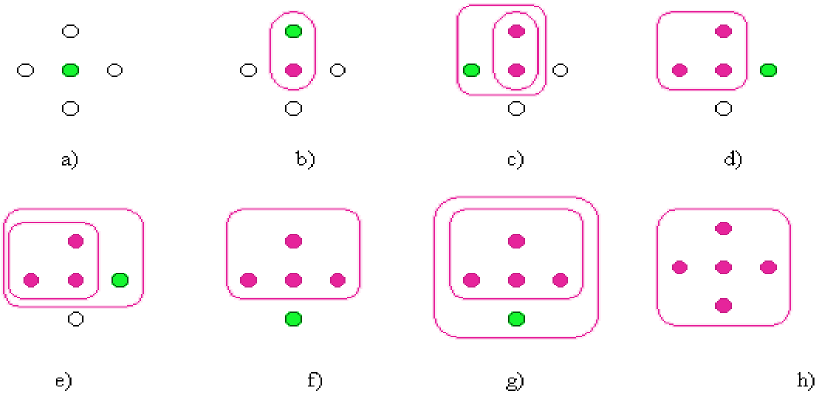


Fig. 5. The stages of forming a cluster

In general, agents trade with each other in the environment called Virtual Market where decisions are made by agent negotiations involving exchange of virtual money. Each agent has a specified sum of monetary units, which it may use to pay joining fees or leaving penalties, if any.

Various microeconomic models can be used to control the clustering process. For example, a charge for joining a cluster may be in accordance with the “club model” (the charge does not depend on the situation, richness of a record or on the number of the club members) or according to the “shareholder model” (the amount depends on the situation). In the latter case, a record may increase its capital entering or leaving the cluster at the right time.

Clusters and records conduct perpetual negotiation, with parties agreeing to meet requested demands or meeting each other half-way, for example, by mutual concessions. As a result of negotiation records may leave clusters and join new ones causing further adjustments, known as a ripple effect. The decentralised decision process combined with microeconomics enables balancing of factors such as the accuracy, cost and time effectiveness.

Multi-agent clustering method enables users to guide the clustering process by specifying clustering rules. Agents representing records or clusters use these rules whenever they have to decide whether particular records and clusters should be matched. For example, the rule of maximising density can be combined with rules on restrictions to the number of clusters to which a record can belong, on the number of records a cluster can contain, on how rich a cluster can be, or what the lifespan of a cluster should be. Various combinations of rules result in different clustering outcomes: few big clusters, a large number of small clusters, very dense clusters, clusters with dispersed records, clusters that are robust, clusters that are dynamic, etc.

The clustering processes for a stable data structure resembles crystallisation processes – first records create elementary structures at the bottom level and then structures themselves get involved in clustering, forming ever more complicated structures, only to stop when available records are exhausted. The clustering outcome is a high-level structure, which is more or less stable, but is modified in real time as new records arrive.

For the task of analysing semantic descriptors of records and clusters, a metrics is introduced, which provides a measure of semantic proximity of descriptors. The metrics is based on ontology. Using heuristic methods distances between pairs of ontology concepts are determined taking into account factors such as belonging to the same parents, having the same attributes and connected by the same relation. The metrics is valid for the comparison of two records but not for comparison of several records because transitivity rule is not valid (if document A is close to document B, and document B is close to document C, we cannot say how close are A and C). The lack of transitivity makes the use of conventional algorithmic methods for the comparison of semantic proximity impossible.

4. Testing

Tests of the proposed method were conducted using documents which were previously manually processed by lawyers of the client company and divided into groups according to their semantic similarity. Templates were previously manually created based on this analysis. Tests consisted of (1) the construction of problem domain ontology based on the full set of documents; (2) creation of semantic descriptors for each document using Magenta semantic analysis engine; (3) clustering of semantic descriptors using Magenta clustering engine; and (4) creating templates for each cluster in semi-automatic mode with the help of the expert who was not involved in the preparation of the test set.

# of Docs	# of Groups* (auto/ manual)	Max hierarchy level (auto / manual)	Average number of docs in a group (auto / manual)	Same and similar groups (%)	Template validity level (%)
11	7 / 9	4/3	2/2	94 %	96 %
43	23/14	5/2	4/3	91 %	90 %
125	54/28	9/3	6/4	87 %	81 %
864	279 / 43	13/3	5/14	88 %	79 %

shows overall number of groups rather that only high-level groups

Table 1. Comparing experts manual test analysis with system result

The testing generated several interesting observations: The smaller the number of documents, the better is the result of manual analysis (which is to be expected). Experts on average prefer to simplify the document structure; therefore the number of hierarchical levels and the number of groups produced in manual mode is considerably lower, groups usually being: “very similar, quite similar, more or less similar, and non-similar”. In all tests one or several groups specified by expert (usually on top levels of hierarchy, i.e. more generalised) were the same as those produced in automatic mode. Differences were caused by the system feature to include a document into every cluster with which the document is semantically similar, while human experts prefer to assign such documents to only one cluster. This difference in the assignment of documents to clusters is the reasons for the increasing divergence between manual and automatic results over time – when there are many documents, human experts have difficulties with identifying and remembering similarities. In groups, which were found both by Magenta engines and human experts, there were many more documents that were missed by experts (approx. 35%), than documents that were wrongly assigned (11%) or missed by clustering engines and found by experts (7%). To summarise testing results, system even in the purely automatic mode, without any human guidance, gave results with approximately 90% agreement in grouping and 80-85% in template building. In practical testing, when Magenta engines were fully implemented and the system was fully deployed in the client’s company, the results were as follows. The processing of approximately 25,000 instances of car insurance contracts, each of 30 pages, which was estimated to require 16 man-years, was done by the system based on the method described in this paper in 40 man-months, i.e. the productivity gain of almost five times.

5. Conclusion

Multi-agent method for semantic analysis and clustering of documents described in this paper has been applied to practical non-trivial problems and achieved results many times better than those accomplished by experts. To the best of our knowledge there are no published applications of either conventional or multi-agent systems that have achieved comparable (or indeed any) results. The success of the Magenta method is based on the following main features: (1) ontological modelling of the problem domain; (2) reformulation of the semantic analysis and clustering problems into problems of the allocation of meaning to words and clusters to records, respectively; (3) distributed creation of solutions in a event-driven, stepwise manner through a process of knowledge-based (rather than data-driven) negotiation (exchange of messages) between autonomous agents rather than by the predetermined algorithmic procedures; (4) creation of a powerful run-time engine capable of supporting communication links and decision making logic for hundreds of thousands agents; (5) implementation of user-friendly ontology editors and user interfaces for users. Authors have applied the method to a variety of practical problems for which there were no known solutions, including semantic search, generation of logistics rules and fault and fraud detection.

References

1 Dumains Susan T., Furnas George W., Landauer Thomas K.: Indexing by Latent Semantic Analysis. Bell Communications Research 435 South St. Morristown, NJ 07960. Richard Rashman: University Of Western Ontario.

2 Douglass R. Cutting, David R. Karger, Jan O. Pedersen, and John W. Tukey: Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections, SIGIR ’92. 1992, 318 – 329

3 Zamir Oren Eli.: A Phrase-Based Method for Grouping Search Engine Results. University of Washington, Department of Science & Engineering.

4 Steinbach, M., Karypis, G., and Kumar, V. (2000).: A Comparison of Document Clustering Techniques, Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 20-23, 2000, Boston, MA, USA.

5 Andreev V., Iwkushkin K., Minakov I., Rzevski G., Skobelev P.: The Constructor of Ontologies for Multi-Agent Systems. In 3rd International Conference ‘Complex Systems: Control and Modelling Problems’, Samara, Russia, September 4-9 2001, 480 – 488.

6 Morin E.: Automatic acquisition of semantic relations between terms from technical corpora. Proc. Of the Fifth Int. Congress on Terminology and Knowledge Engineering (TKE-99), TermNet-Verlag, Vienna

7 Faure D, Poibeau T.: First experiments of using semantic knowledge learned by ASIUM for information extraction task using INTEX. In: S. Staab, A. Maedche, C. Nedellec, P. Wiemer-Hastings (eds.), Proceedings of the Workshop on Ontology Learning, 14th European Conference on Artificial Intelligence ECAI’00, Berlin, Germany. 2000.

8 Andreev V., Iwkushkin K., Karyagin D., Minakov I., Rzevski G., Skobelev, P., Tomin M.: Development of the Multi-Agent System for Text Understanding. In 3rd International Conference ‘Complex Systems: Control and Modelling Problems’. Samara, Russia, September 4-9 2001, 489 – 495.

9 I Minakov, G Rzevski, P Skobelev: Automated Text Analysis // Patent Application No. 305634, UK, 2004.

10 I Minakov, G Rzevski, P Skobelev: Data Mining // Patent Application No. 0403145.6, UK, 2004.

11 Rzevski G., Petr Skobelev, Igor Minakov and Semen Volman “Dynamic Pattern Discovery using Multi-Agent Technology”. Proceedings of the 6th WSEAS International Conference on Telecommunications and Informatics (TELE_INFO ’07), Dallas, Texas, USA, March 22-24, 2007, 75-81.

12 Skobelev P.O.: Holonic Systems Simulation // Proc. of the 2nd International Conference “Complex Systems: Control and Modelling Problems”, Samara, June 20-23, 2000, 73-79.

Published: Minakov I., Rzevski G., Skobelev Petr, Volman S. Creating Contract Templates for Car Insurance Using Multi-Agent Based Text Understanding and Clustering // HoloMAS 2007. Volume 4659 LNAI, 2007, Pages 361-370.