Group Recommendation System for User-Centric Support in Virtual Logistic Hub

Architecture and Major Components

Alexander Smirnov, Nikolay Shilov
St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences
St. Petersburg, Russia
{smir; nick}@iias.spb.su

***Abstract*—Currently, the society becomes more and more knowledge-intensive when a level of collaboration of different groups of people and institutes increases dramatically. One of the possible ways to assist to a group of users is collaborative recommendation systems. These systems have to recommend some solutions (related to products, technologies, tools, material and business models) based on user group requirements, preferences and willingness to compromise and to be pro-active. The paper proposes an approach to developing a group recommendation system for virtual logistic hub based on such technologies as user and group profiling, context management, decision mining. The system allows accumulation of knowledge about user actions and decisions and compromising between group and individual preferences. Proposed approach enables formulation of recommendations for users of the same group anticipating their possible further actions and decisions.**

Keywords-collaborative recommendation system; group profiling; context management; decision mining.

I. INTRODUCTION

Small and Medium businesses (SMEs) and personal travel via cars, buses and trains is usually (and reasonably) done within the radius of 450-500 kilometers. The distance between St. Petersburg, Russia and Helsinki, Finland together with nearby cities (Imatra, Lappeenranta, Kotka, Vyborg) falls into this radius. Taking into account available airports in Helsinki, Lappeenranta, and St. Petersburg as well as ferries in Helsinki, Kotka, and St. Petersburg, this region constitutes a universal hub for travelling all around the world.

In order for this hub to function, an efficient transportation system within the region has to be formed. However, today the travelling in the region is complicated due to a number of reasons, e.g., unpredictable situation at border crossing, unknown traffic condition on the roads, isolation of train, bus, and airplane schedules. The proposed approach is aimed at support of dynamic configuration of virtual multimodal logistics networks based on user requirements and preferences. The main idea is to develop models and methods that would enable ad-hoc configuration of resources for multimodal logistics. They are planned to be based on dynamic optimization of the route and transportation means as well as to take into account user preferences together with unexpected and unexpressed needs (on the basis of the profiling technology).

The small business and personal travelling is characterized by the following features: non-regular, not expensive, and safe. As a result, the proposed approach assumes developing a group recommendation system for ad hoc generation of travel plans for the region (the South of Finland and St. Petersburg region) taking into account the current situation on the roads and border crossings, fuel management aspects, travel time and distance. The increase of travelling will be a significant step towards development of the integrated economic zone in the Region.

Until recently, the most recommendation systems operated in the 2-dimentional space “user-product”. They did not take into account the context information, which, in most applications can be critical. As a result, there was a need in development of group recommendation systems based not only on previously made decisions but also on the contexts of situations in which the decisions were made. This gave a rise to development of context-driven collaborative algorithms of recommendation generation since their usage would significantly increase the quality and speed of decision making.

Besides, the proposed general framework will be a channel for collecting user’s feedback, preferences and demands for new services that users cannot find in the Region or quality of which shall be improved. What is important is that not only the problem is identified, but in most cases immediate hints/suggestions can be provided regarding what shall be done to better serve users’ needs.

The framework will also significantly benefit to the ecological situation in the region via reducing not necessary transportation and waiting time for border crossing. In accordance with Global GHG Abatement Cost Curve v 2.0 [1] in the travelling sector the carbon emission can be significantly decreased via more efficient route planning, driving less, switching from car to rail, bus, cycle, etc. As a result, evolving of flexible energy and eco-efficient logistics systems can be considered as one of the significant steps towards the knowledge-based low carbon economy.

The paper is structured as follows. The next section introduces the virtual logistic hub. It is followed by the description of the approach. Then, the group recommendation system architecture is proposed. The knowledge representation formalism used in the developed

 ISBN: 978-1-61208-224-0

approach is presented in sec. V. Sec. VI presents the user clustering algorithm, followed by the description of how the common preferences/interests are identified (sec. VII). The main results are summarized in the Conclusion.

II. VIRTUAL LOGISTIC HUB

The idea of virtual logistic hub has already been mentioned in the literature (though it could have a different name, e.g., “e-Hub” [2]), but it is still devoted very little attention in the research community. For example, [3] and [4] consider the virtual logistic hub from organizational and political points of view. Generally, virtual logistic hub represents a virtual collaboration space for two types of members: (i) transportation providers (who actually moves the passengers or cargo), and (ii) service providers (who provides additional services, e.g., sea port, border crossing authorities, etc.). These providers can potentially collaborate in order to increase the efficiency of the logistic network (solid lines in Figure 1), however, it is not always the case. The major idea of the virtual logistic hub is to arrange transportation based on the available schedules and capabilities of transportation and service providers, current and foreseen availability and occupancy of the transportation means and services (“dash-dot” lines in Figure 1). In this case, even though the schedules and actions of different members are not coordinated, the virtual logistic hub will be able to find the most feasible transportation schedule depending on the current situation and its likely future development. For the end-user (travelers or cargo owners), all this is hidden “under the hood”, and only the final transportation schedule is seen (solid lines in Figure 1).

III. APPROACH

Figure 2 represents the generic scheme of the approach. The main idea of the approach is to represent the logistics system members by sets of services provided by them. This makes it possible to replace the configuration of the logistics system with that of distributed services. For the purpose of semantic interoperability, the services are represented by Web-services using the common notation described by a common ontology. The agreement between the resources and the ontology is expressed through alignment of the descriptions of the services modeling the resource functionalities and the ontology. As a result of the alignment operation the services get provided with semantics. The operation of the alignment is supported by a tool that identifies semantically similar words in the Web-service descriptions and the ontology. In the proposed approach the formalism of Object-Oriented Constraint Networks (OOCN) is used (its detailed description can be found in [20]) for knowledge representation in the ontology (see sec. V).

Depending on the problem considered, the relevant part of the ontology is selected forming an abstract context. The abstract context is an ontology-based model embedding the specification of problems to be solved. It is created by core services incorporated in the environment. When the abstract context is filled with values from the sources, an operational context (formalized description of the current situation) is built. The operational context is an instantiated abstract context and the real-time picture of the current situation. Producing the operational context is one of the purposes of resource configuration. Since the resources are represented by sets of services, the configuration of the resources is replaced with that between the appropriate services. Besides the operational context producing, the services are purposed

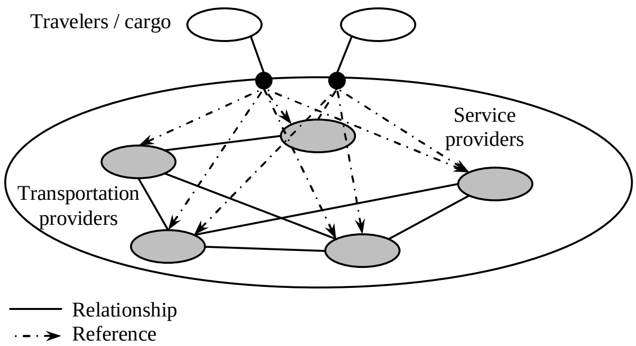

Figure 1. Generic scheme of the virtual logistic hub

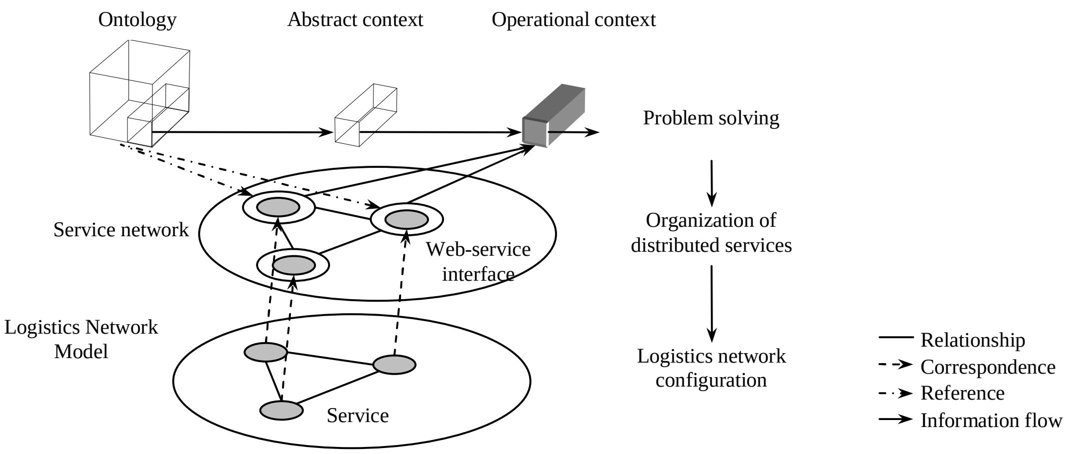

Figure 2. Generic scheme of the approach

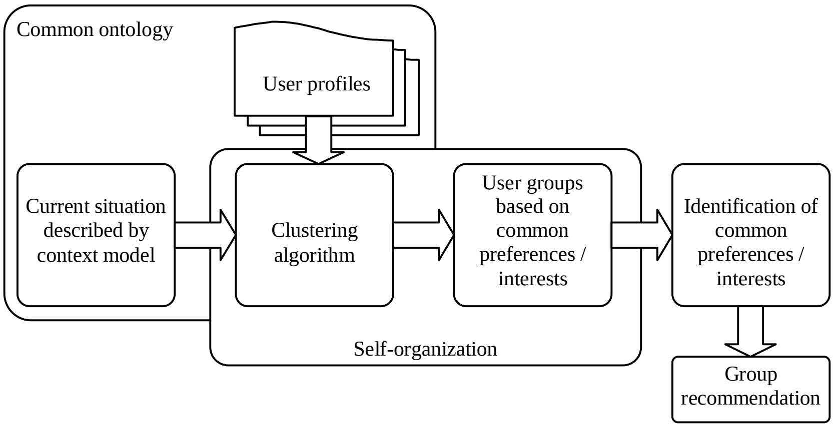

Figure 3. Group recommendation system architecture

to solve problems specified in the abstract context and to get the resources to take part in logistics plan. Due to the usage of the OOCN formalism the operational context represents the constraint satisfaction problem that is used during organisation of services for a particular task.

It can be guessed that for each particular situation there can be a large amount of feasible solutions for the users to choose from (e.g., the fastest transportation, the least amount of transfers, sightseeing routes, etc.). As a result, the paper proposes to build such a system as a group recommendation system that learns user preferences and recommends solutions, which better meet those preferences.

IV. GROUP RECOMMENDATION SYSTEM ARCHITECURE

Generation of feasible transportation plans taking account explicit and tacit preferences requires strong IT-based support of decision making so that the preferences from multiple users could be taken into account satisfying both the individual and the group [5]. Group recommendation systems are aimed to solve this problem.

Recommendation / recommending / recommender systems have been widely used in the Internet for suggesting products, activities, etc. for a single user considering his/her interests and tastes [6], in various business applications (e.g., [7], [8]) as well as in product development (e.g., [9], [10]). Group recommendation is complicated by the necessity to take into account not only personal interests but to compromise between the group interests and interests of the individuals of this group.

There are two major types of recommending systems: (i) content-based (recommendations are based on previous user choices), and (ii) collaborative filtering (recommendations are based on previous choices of users with similar interests). The second type is preferable for the domains with larger amounts of users and smaller activity histories of each user, which is the case for the logistics hub.

In literature (e.g., [11], [12]) the architecture of the collaborative filtering recommending system is proposed based on three components: (i) profile feature extraction from individual profiles, (ii) classification engine for user clustering based on their preferences (e.g., [13]), and (iii) final recommendation based on the generated groups. The core of such system is a clustering algorithms capable to continuously improve group structure based on incoming information enables for self-organization of groups [14].

The proposed group recommendation system architecture for logistics hub is presented in Figure 3. It is centralized around the user clustering algorithm [15] originating from the decision mining area [16]-[18]. The proposed clustering algorithm is based on the information from user profiles. The user profiles contain information about users including their preferences, interests and activity history (a detailed description of the profile can be found in [19]). Besides, in order for the clustering algorithm to be more precise, this information is supplied in the context of the current situation (including current user task, time pressure and other parameters). The semantic interoperability between the profile and the context is supported by the common ontology.

The user profiles are considered to be dynamic and, hence, the updated information is supplied to the algorithm from time to time. As a result, the algorithm can run as updated information is received and update user groups. Hence, it can be said that the groups self-organize in accordance with the changes in the user profiles and context information.

When groups are generated the common preferences / interests (e.g., the fastest transportation, the least amount of transfers, sightseeing routes, etc.) of the groups are identified based on the results of the clustering algorithm. These preferences are then generalized and analyzed in order to produce group recommendations.

Usage of an appropriate knowledge representation formalism is one of the keys to development of an efficient clustering algorithm.

V. KNOWLEDGE REPRESENTATION FORMALISM

Since the user profiles and the current situation context are analyzed jointly, it is reasonable to use the same formalism and terminology for their representation. In the proposed approach the formalism of Object-Oriented Constraint Networks (OOCN) is used (its detailed description can be found in [20]) for knowledge representation in the ontology. It provides primitives for modelling classes, class hierarchies and other class structures, class attributes, attribute inheritance, attribute ranges, and functional dependencies.

According to this formalism the ontology A is represented by sets of classes, class attributes, attribute domains, and constraints:

$A = \langle O, Q, D, C \rangle$, where

O – a set of *object classes* ("classes")

Q – a set of class *attributes* ("attributes");

D – a set of attribute *domains* ("domains");

C – a set of *constraints* used to model relationships.

The set of constraints includes six types of constraints for modelling different relationships:

C_1 – (class, attribute, domain) relation used to model triple of classes, attributes pertinent to them, and restrictions on the attribute value ranges;

C_2 – taxonomical ("is-a") and hierarchical ("part-of") relations used to model class taxonomy and class hierarchy respectively;

C_3 – classes compatibility used to model condition if two or more instances can be parts of the same class;

C_4 – associative relationships used to model any relations and axioms of external ontologies neglected by the internal formalism;

C_5 – class cardinality restriction used to define how many subclasses the class can have;

C_6 – functional relations used to model functions and equations.

Such representation of knowledge can be interpreted as a constraint satisfaction task and used by a constraint satisfaction / propagation engines for reasoning and optimization.

Below, some example constraints are given:

- an attribute *costs* (q_1) belongs to a class *ride* (o_1): $c^{I}_{1} = (o_1, q_1)$;
- the attribute *costs* (q_1) belonging to the class *ride* (o_1) is a real number: $c^{II}_{1} = (o_1, q_1, \mathrm{R})$;
- a class *cargo* (o_2) is compatible with a class *truck* (o_3): $c^{III}_{1} = (\{o_2, o_3\}, True)$;
- an instance of the class *ride* (o_1) can be a part of an instance of a class *travel* (o_4): $c^{IV}_{1} = \langle o_1, o_4, 1\rangle$;
- the *truck* (o_3) is a *resource* (o_5): $c^{\mathrm{IV}}_{1} = \langle o_3, o_5, 0\rangle$;
- an instance of the class *cargo* (o_2) can be connected to an instance of the class *truck* (o_3): $c^{V}_{1} = (o_2, o_3)$;
- the value of the attribute *cost* (q_1) of an instance of the class *travel* (o_4) depends on the values of the attribute *cost* (q_1) of instances of the class *ride* (o_1) connected to that instance of the class *travel* and on the number of such instances: $c^{VI}_{1} = f(\{o_1\}, \{(o_4, q_1), (o_1, q_1)\})$.

VI. USER CLUSTERING ALGORITHM

Due to the specific of the tasks in the considered domain the implemented algorithm (adapted from [15]) of user clustering is based on analysing user preferences and solutions selected by users and has the following steps:

1. Preliminary linguistic analysis of preferences (tokenisation, spelling and stemming).
2. Extract words/phrases from the preferences and solutions (text processing).
3. Find ontology elements occurring in the extracted words and phrases.
4. Construct weighted graph consisting of ontology classes and attributes, and users. Weights of arcs are calculated on the basis of (i) similarity metrics (i.e. they are different for different user solutions) and (ii) taxonomic relations in the ontology.
5. Construct weighted graph consisting of users (when classes and attributes are removed, arcs' weights are recalculated).
6. Cluster users graph.

Finding ontology elements occurring in the extracted words and phrases is done in two ways: (i) via syntactic similarity, and (ii) via semantic similarity.

The syntactic similarity is calculated via the algorithm of fuzzy string comparison similar to the well-known Jaccard index [21]. It calculates occurrence of substrings of one string in the other string. For example, string "motor" has 5 different substrings (m, o, t, r, mo) contained in the string "mortar". The total number of different substrings in "motor" is 13 (m, o, t, r; mo, ot, to, or; mot, oto, tor; moto, otor). The resulting similarity of the string "motor" to the string "mortar" is 5/13 or 38%.

The semantic similarity (or distance) is based on the machine-readable dictionary Wiktionary [22]. The ontology is represented as a semantic network where names of classes and properties constitute nodes of the network. The nodes corresponding to the ontology elements are linked to nodes representing their synonyms and associated words as this is given in the machine-readable dictionary. The links between the nodes are labelled by the weights of relations specified between the concepts represented by these nodes in the machine-readable dictionary. Weight w of a relation specified between two concepts t_i and t_j is assigned as 0,5 if t_i and t_j are synonyms; 0,3 if t_i and t_j are associated words; and $\propto$ if t_i and t_j are the same words. The nodes representing extracted words and phrases are checked for their similarity to nodes representing ontology elements. As a measure of similarity semantic distance *Dist* is used:

$$Dist(t_j, t_j) = 1 / \left(\sum_S \prod_k w_k \right),$$

where S is a set of paths from t_i to t_j, formed by any number of links that connect t_i and t_j passing through any number of nodes (k).

For example, let us suppose that the set of words came of parsing the profile comprises two words: *trip* and *lorry*. An illustrative piece of the semantic network built based on this table and is represented in Figure 4. The Figure illustrates three names for classes and attributes in the ontology corresponding to the extracted words: *Trip*, *Ship*, and *Truck*. The semantic distances are as follows:

$Dist(trip, trip) = 1 / \propto = 0$

$Dist(lorry, ship) = 1 / (0.5*0.3 + 0.3*0.5) = 3,33$

$Dist(lorry, truck) = 1 / (0.5*0.3 + 0.3*0.3*0.3 + 0,5) = 1.48$

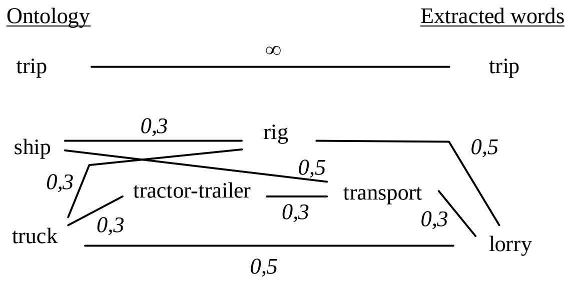

Figure 4. A piece of semantic network relevant to WSDL-attribute "Accident point".

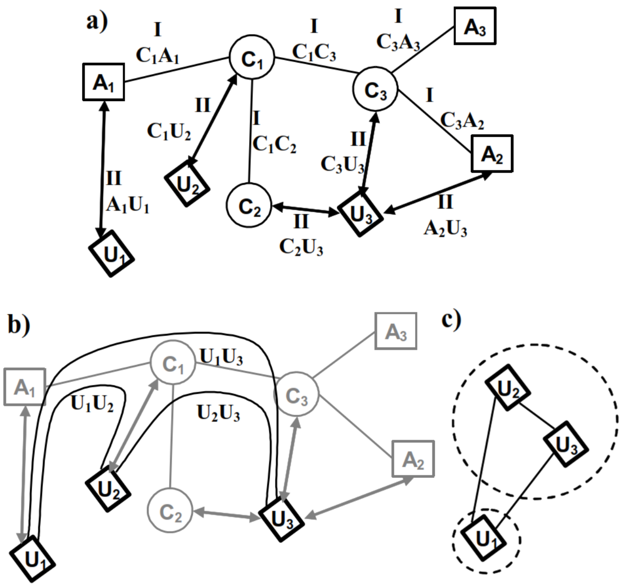

Figure 5. Weighted user – ontology graph and user clustering procedure

It can be seen that for the distance between the concepts *lorry* and *truck* is much shorter than between the concepts *lorry* and *ship*. So, the class *truck* is aligned to the concept *lorry*.

For the clustering procedure, a weighted user – ontology graph is considered. It contains three types of nodes: C – classes from the ontology, A – their attributes, and U – users.

The graph consists of two types of arcs. The first type of arcs I (CA, CC) is defined by the taxonomy of classes and attributes in the ontology. The second type of arcs II (CU, AU) is defined by relations between user solutions and classes/attributes (Figure 5a).

Weights of arc between nodes corresponding to classes and users CU_{weight} and corresponding to attributes and users AU_{weight} are defined via the similarity CU_{sim} and AU_{sim} of the class or attribute (calculated via the fuzzy string comparison algorithm described above). The similarity is a property of relations between class – user/solution or attribute – user/solution. Weights of arcs are defined as follows: $CU_{weignt} = 1 - CU_{sim}$; $AU_{weight} = 1 – AU_{sim}$.

Arcs CA and CC tying together classes and attributes via taxonomic relations (defined by ontology relations class-class, class-attribute) have CA_{weight}, $CC_{weight} \in (\varepsilon, 1)$ defined empirically. CC_{weight} means arcs' weight of linked classes in the ontology. CA_{weight} – arcs' weight of linked attributes and classes.

Since users are represented by their solutions, based on this graph the solutions and weight consequently users are clustered on the basis of the lowest weights of connecting arcs. This is performed in the following sequence. First, the shortest routes between users are calculated (Figure 5b). E.g., weight of the arc U_1U_2 will be calculated as follows: $U_1U_{2\ weight} = A_1U_{1\ weight} + C_1A_{1\ weight} + C_1U_{2\ weight}$; weight of the arc U_2U_3 can be calculated in 3 ways, it is considered in Figure 2b that $U_2U_{3\ weight} = C_1U_{2\ weight} + C_1C_{3\ weight} + C_3U_{3\ weight}$ is the shortest one; etc. Based on the calculated weights a new graph consisting of the users only is built (cf. Figure 2c). The value of the parameter D_{max} is set empirically. Assuming that $U_1U_{2\ weight} > D_{max}$, $U_1U_{3\ weight} > D_{max}$, and $U_2U_{3\ weight} < D_{max}$, two clusters can be identified: the first cluster includes users U_2 and U_3, and the second one includes customer U_1 (dashed circles in Figure 5c).

The algorithm can run as updated information is received and update user groups thus providing for self-organizations of user groups in accordance with the changes in the user profiles and context information.

The developed ontology-based clustering algorithm has the following advantages compared to other clustering techniques: (i) *domain-specific knowledge filter* using the ontology; (ii) *natural language processing*; (iii) *term extraction*, such as ontology classes and attributes, units of measures (e.g., "km" and "hrs") can be extracted from the user preferences.

VII. IDENTIFICATION OF COMMON PREFERENCES/INTERESTS AND GROUP RECOMMENDATIONS

User preferences consist of attributes (properties) and/or their values, classes (problem types), relationships (problem structure) and/or optimization criteria that are usually preferred or avoided by the user. The preference revealing can be interpreted as identification of *patterns of the solution selection* (decision) by a user from a generated set of solutions. The ability to automatically identify patterns of the solution selection allows to sort the set of solutions, so that the most relevant (to user needs) solutions would be in the top of the list of solutions presented to the user.

Currently, three major tasks of identification of user preferences can be selected:

1. Identification of *user preferences based on solutions generated for the same context*. In this case, the problem structure is always the same, however its parameters may differ.
2. Identification of user *preferences based on solutions generated for different contexts*. This task will be more complex then the first one since structures of the problem will be different.
3. Identification of *user preferences in terms of optimization parameters*. This task will try to identify if a user tends to select solutions with

Copyright (c) IARIA, 2012. ISBN: 978-1-61208-224-0

minimal or maximal values of certain parameters (e.g., time minimization) or their aggregation.

Based on the clusters built, the user preferences can be identified as common preferences of the users grouped into the clusters.

VIII. CONCLUSION

The paper presents an approach to development of group recommendation system for virtual logistic hub. Virtual logistic hub performs ad-hoc transportation scheduling based on the available schedules, current and foreseen availability and occupancy of the transportation means and services even though they do not cooperate with each other. The approach is based on application of such technologies as user and group profiling, context management, decision mining. It enables for self-organization of user groups in accordance with changing user profiles and the current situation context.

Presented research is at an early development stage. The future work is aimed at implementation of the proposed system in a limited domain for validation of its applicability and efficiency.

ACKNOWLEDGMENT

Some parts of the research were carried out under projects funded by grants # 11-07-00045-a, # 12-07-00298-a, # 12-07-00302-a of the Russian Foundation for Basic Research, project 2.2 of the Nano- & Information Technologies Branch of the Russian Academy of Sciences, and project of the research program "Information, control, and intelligent technologies & systems" of the Russian Academy of Sciences.

REFERENCES

[1] Global GHG Abatement Cost Curve v 2.0, 2009, https://solutions.mckinsey.com/ClimateDesk/default.aspx. Retrieved: June, 2011.

[2] E. Chang, T. Dillon, W. Gardner, A. Talevski, R. Rajugan and T. Kapnoullas, "A Virtual Logistics Network and an E-hub as a Competitive Approach for Small to Medium Size Companies," Web and Communication Technologies and Internet-Related Social Issues – HIS 2003, Lecture Notes in Computer Science, vol. 2713, 2003, pp. 167-168.

[3] Working Group on Logistics, "Developing Singapore into a Global Integrated Logistic Hub," Report, 2002. http://app.mti.gov.sg/data/pages/507/doc/ERC_SVS_LOG_MainReport.pdf. Retrieved: February, 2012.

[4] E. Sweeney, "Supply Chain Management in Ireland:the Future," Logistics Solutions, the Journal of the National Institute for Transport and Logistics, Vol. 5, No. 3, pp. 14-16, June 2002.

[5] K. McCarthy, M. Salamo, L. Coyole, L. McGinty, B. Smyth, P. Nixon, "Group Recommender Systems: A Critiquing Based Approach," IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces, 2006, pp. 267-269.

[6] I. Garcia, L. Sebastia, E. Onaindia, C. Guzman, "Group Recommender System for Tourist Activities," EC-Web 2009: Proceedings of E-Commerce and Web Technologies, The 10th International Conference (2009), Springer, LNCS 5692, 2009, pp. 26-37.

[7] T. Hornung, A. Koschmider, A. Oberweis, A. "A Recommender System for Business Process Models," In WITS'09: Proceedings of the 17th Annual Workshop on Information Technologies & Systems, 2009. Electronic resource. http://ssrn.com/abstract=1328244. Retrieved: February, 2012.

[8] L. Zhena, G. Q. Huangb, Z. Jiang, "Recommender system based on workflow," Decision Support Systems, Elsevier, vol. 48, no. 1, 2009, pp. 237-245.

[9] S. K. Moon, T. W. Simpson, S. R. T. Kumara, "An agent-based recommender system for developing customized families of products," Journal of Intelligent Manufacturing, Springer, vol. 20, no. 6, 2009, pp. 649-659.

[10] Y.-J. Chen, Y.-M. Chen, M.-S. Wu, "'An expert recommendation system for product empirical knowledge consultation," ICCSIT2010: The 3rd IEEE International Conference on Computer Science and Information Technology, IEEE, 2010, pp. 23-27.

[11] E.-A. Baatarjav, S. Phithakkitnukoon, R. Dantu, R. "Group Recommendation System for Facebook," OTM 2008: Proceedings of On the Move to Meaningful Internet Systems Workshop (2008), Springer, LNCS 5333, 2008, pp. 211-219.

[12] S. E. Middleton, D. De Roure, N. R. Shadbolt, "Ontology-Based Recommender Systems," Handbook on Ontologies (Ed. By S. Staab, R. Studer), Springer, 2003, pp. 477-498.

[13] H. C. Romesburg, "Cluster Analysis for Researchers," Lulu Press, California, 2004, 344 p.

[14] G. W. Flake, S. Lawrence, C. L. Giles, F. Coetzee, "Self-Organization and identification of Web Communities," IEEE Computer, vol. 35, no. 3, 2002, pp. 66-71.

[15] A. Smirnov, M. Pashkin, N. Chilov, "Personalized Customer Service Management for Networked Enterprises," ICE 2005: Proceedings of the 11th International Conference on Concurrent Enterprising, 2005, pp. 295-302.

[16] A. Smirnov, M. Pashkin, T. Levashova, A. Kashevnik, N. Shilov, "Context-Driven Decision Mining," Encyclopedia of Data Warehousing and Mining. Hershey (Ed. by J. Wang), New York, Information Science Preference, Second Edition. vol. 1., 2008, pp. 320 – 327.

[17] A. Rozinat, W. M. P. van der Aalst, "Decision Mining in Business Processes," BPM Center Report no. BPM-06-10, 2006.

[18] P. Petrusel, D. Mican, "Mining Decision Activity Logs," BIS2010: Proceedings of Business Information Systems Workshops, Springer, LNBIP 57, 2010, pp. 67–79.

[19] A. Smirnov, T. Levashova, A. Kashevnik, N. Shilov, "Profile-based self-organization for PLM: approach and technological framework," PLM 2009: Proceedings of the 6th International Conferecne on Product Lifecycle Management, Electronic proceedings, 2009.

[20] A. Smirnov, T. Levashova, N. Shilov, "Semantic-oriented support of interoperability between production information systems," International Journal of Product Development, Inderscience Enterprises Ltd., vol. 4, no. 3/4, 2007, pp. 225-240.

[21] P.-N. Tan, M. Steinbach, V. Kumar, "Introduction to Data Mining," Addison Wesley, 2005, 769p.

[22] Wiktionary, 2012. http://www.wiktionary.org. Retrieved: February, 2012.