IOP Conf. Series: Earth and Environmental Science **776** (2021) 012007 doi:10.1088/1755-1315/776/1/012007

Big data aggregation in disasters risk management systems

Oleksiy Kovalenko[1] and Dimiter Velev[2]

1 IMMSP NAS of Ukraine, Kyiv, Ukraine
2 University of National and World Economy, Sofia, Bulgaria
E-mail: koval@immsp.kiev.ua and dgvelev@unwe.bg

Abstract. In the article the approach to aggregation of Big Data in the disasters risk management systems with using knowledge models is discussed. The basic concept is classifying and structuring the information received and stored in disasters management system using specific attributes of information, represented in the model of knowledge of disasters. We study the hierarchy of disasters risk management system and specificity of storing knowledge of disaster based on ontologies. Situation awareness is one of important elements in the complex problem of situation management. Disasters awareness is a phase of situational, or emergency, or crisis management in the case of disasters management.

1. Introduction

By United Nations Office for Disaster Risk Reduction (UNISDR) definition disaster is '*a serious disruption of the functioning of a community or a society at any scale due to hazardous events interacting with conditions of exposure, vulnerability and capacity, leading to one or more of the following: human, material, economic and environmental losses and impacts*.' [1].

Disaster risk management according to the terminology of UNISDR [2] '*is the application of disaster risk* reduction *policies and strategies to prevent new disaster risk, reduce existing disaster risk and manage residual risk, contributing to the strengthening of resilience and reduction of disaster losses. Disaster risk management actions can be distinguished between prospective disaster risk management, corrective disaster risk management and compensatory disaster risk management, also called residual risk management.*' The first phase of disaster risk management is disasters awareness.

Disasters are results of evolution of situations concerned with different environment natural, technogene and human processes. We can say that disaster is a catastrophic situation caused by different factors. Situation, in general sense, defined as all of the facts, conditions, and events that affect someone or something at a particular time and in a particular place [3].

Situation awareness is one of important elements in the complex problem of situation management [4, 5]. A important kind of situation awareness is disasters awareness in disasters management [6, 7]. Correct disasters awareness is a first step to prevention and elimination of their consequences in context of expression "forewarned is forearmed". In its turn, disasters awareness is a phase of situational, or emergency, or crisis management in the case of disasters management.

Varieties of disasters and their impact on human society calls necessity of processing huge volumes of information called Big Data. Important issue of disasters awareness and disasters management is adoption and assessing of big data volume about disasters. Furthermore, data about disasters requires classification and aggregation by types. Agent-based approach can be used to solve this problem.

2. The analysis of the current state of research and publications

By definition '*data aggregation is a type of data and information mining process where data is searched, gathered and presented in a report-based, summarized format to achieve specific business* *objectives or processes and/or conduct human analysis*' [8]. Data aggregation for disasters awareness in modern context involves information technologies such as data mining, big data processing, data analysis and visualization, knowledge management etc. Therefore, during data aggregation there are raises following problems: selection of data source(s); development of data aggregation strategy; quality assessment of data; data enrichment; data refining; data warehousing; data visualization.

There are different methods for resolving each of these problems, and choice of concrete method is complex task concerned with the use of knowledge from appropriate domain. Therefore, integrated implementation of technologies based on data consolidation methods in disasters management process needs appropriate means, in particular, use of program agents having appropriate knowledge.

Fundamental study [9] depicts variety methods and problems concerning with data mining. Means and technologies for working with big data are represented in works [10, 11]. Knowledge management problems are outlined in works [12-14]. Data analysis and visualization are broad fields of activities that were described, particularly in works [15-17]. Using program agents for information processing implies definition of agents' model, behavior, and coordination mechanism for cooperative working. Issues of agent technologies was presented in works [18-20]. Some conceptual and architectural aspects of creation intelligent systems for disasters management, without taking into account big data processing, were considered in works [5, 21].

3. Statement of the problem

Disasters awareness in first needs classification and adoption of information about disasters. The key aspect of adequate awareness of any situation is accuracy, purity, and convenience of perception of information. These properties can be achieved based on the knowledge about data and data processing methods at all stages of data science process [14, 15]. Therefore, creation of means for appropriate data preparation for solving specific tasks from different domains is actual problem. One of this domain is disasters risk management.

4. The main problem

Necessity of prompt situation awareness in disasters risk management raises the problem of fast preparation of disasters data. Variety and huge volumes of information during data processing requires the use of knowledge based approach. This approach can be realized on the base of program agents' community systems. The main problems of creation agents' communities for big data aggregation is development of domain specific knowledge base, allocation and assignment roles, supporting communication and coordination mechanisms between agents.

5. Agents communities knowledge for big data aggregation

The first step for development of agents' community is creation of domain specific knowledge base (KB). This KB can be created on the basis of ontology. The fundamentals of KB lays the metaontology with key concept things of disasters problems domain. Generic concepts of this metaontology are mined from classification of disasters. Other concepts of metaontology for agents' knowledge concerns with sources, formats and specifics of big data for disasters awareness, situational contexts of disasters, utilitarian usage of big data, programs models and platforms for big data processing.

5.1. Disaster big data and knowledge

Disaster risk management is based on knowledge which main source is the results of big data processing. This knowledge is of two types – knowledge about disaster situation for its comprehension and knowledge about action in accordance to awareness of situation.

What is knowledge? From knowing to knowledge. Oxford dictionary defines *knowing* as the state of being aware or informed; *knowledge* as facts, information, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject [22].

Disasters big data metaontology includes such ontologies:

- Disaster ontology
- Big data sources ontology
- Big data types ontology

IOP Conf. Series: Earth and Environmental Science **776** (2021) 012007 doi:10.1088/1755-1315/776/1/012007

- Big data platforms ontology
- Agents ontology

Disasters big data metaontology is outlined on fig. 1.

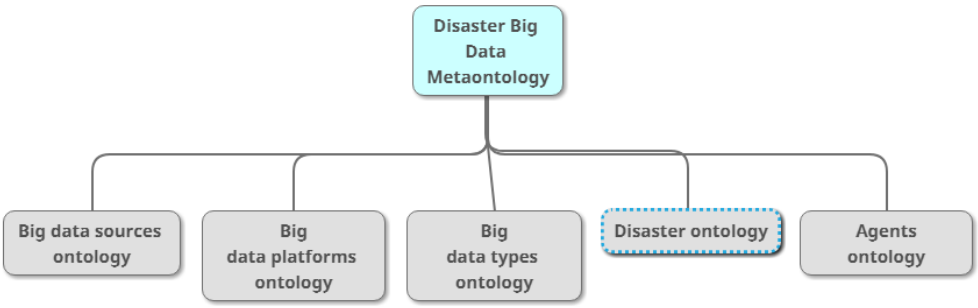

Figure 1. Disasters big data metaontology

Big data algorithms selected by using knowledge. Hence, ontology of knowledge domain is important for the choosing big data process. Ontology formally represents knowledge of specific problem domain as a hierarchy of concepts within this problem domain with the use of domain vocabulary to denote the types, properties and interrelationships of domain concepts.

Disaster classification

Because disasters are results of one or more hazardous events, then detailed classification of disasters and their consequences is not possible. The International Disaster Database EM-DAT (http://www.emdat.be/classification) use five-degree classification of disasters by 1) disasters group, 2) disasters subgroup, 3) disasters main type, 4) disasters sub-type, and 5) disasters sub-sub-type degrees. So, the data about each type of disasters have own peculiarity and volume. Disasters classification consists of two groups of disasters:

- Natural;
- Technological.

Natural disasters are classified by subgroups:

- Geophysical;
- Meteorological;
- Hydrological;
- Climatological;
- Biological;
- Extraterresial.

Technological disasters concern with:

- Industrial accidents;
- Transport accidents;
- Miscellaneous accidents.

Attributes of disasters are defined their type.

Big Data Sources for Disasters Awareness

Data and information about disasters comes from:

- Archives;
- Enterprise sources;
- Transactional sources;
- Social media;
- Public sources;
- Different activities.

Big data categories, formats and structure

Another category of ontology is mission of big data for utilitarian use. Big data categories is differs by utilitarian use for:

- Interaction;

- Description;
- Behavioral characterization;
- Attitudinal expression.

Formats of big data are
- Simple (arithmetic, symbolic)
- Complex (structured simple, audial, visual).

By the structure big data may be:
- Structured;
- Semistructured;
- Quasistructured;
- Unstructured.

Structuring of big data is depicted on fig.2.

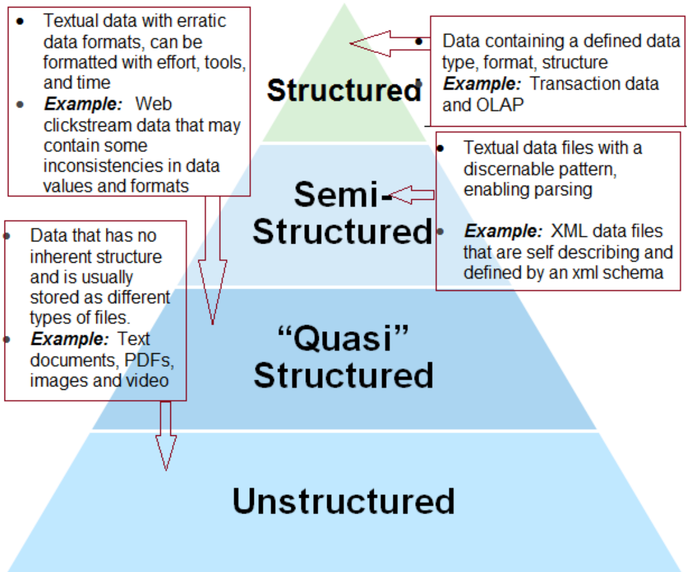

Figure 2. Structuring of big data [23]

Structured Data containing a defined data type, format, structure, for example transaction data and OLAP, relational database, spreadsheets. Semi-Structured Data mainly textual data files with a discernable pattern, enabling parsing, for example, XML data files that are self-describing and defined by an xml schema. Quasi-Structured Data mainly textual data with erratic data formats can be formatted with effort, tools, and time, for example, Web clickstream data that may contain some inconsistencies in data values and formats. Unstructured Data is a data that has no inherent structure and is usually stored as different types of files, for example, text documents, PDFs, images and video.

Structuring of information can be made on different layers of presentation. So, structured plain types of data are aggregated and composed in complex types those can be referred to appropriate type of structuring. This schema is essentially decision tree for data classification.

Because big data have different resources one has different structure described appropriate metadata. According to data origin metadata are describe common and/or specific structure of data. Most common metadata vocabulary for different resources and objects is Dublin Core [24] endorsed and used in different standards to define specific data entities. For example, common metadata standards are IETF RFC 5013, ISO 15836 and NISO Z39.85, and specific metadata are DCMI, GILS, EADS, vCard, FOAF, GRIB, VICAR, NewsXML, QuDEx [25], etc.

Adequate processing of big data is based on using appropriate metadata of problem domain. Metadata definitions are stored and maintained in a controlled method in a metadata registry. Hence, each organization or institution has its own metadata registry. Enough comprehensive metadata registries have US Environment Protection Agency (USEPA). In particular, USEPA Reusable Component

Services (RCS) Registry contains references to a lot of useful means and tools for data processing. For unifying of big data handling and processing it is necessary to develop appropriate data model. For this purpose are used metadata and analytic (assessment, modelling, simulation, decision support and other) frameworks [26].
One of the risk management framework referenced in RCS Registry is Adaptive Risk Assessment Modeling System based on Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES) by Pacific Northwest National Laboratory [27]. For RAD of applications there are many frameworks. One of those is CUBA platform. For big data governance is useful Apache Atlas framework.

5.2. Method for processing of big data, determined by its context.
Specifics of big data is characterized by “three V’s”, “five V’s”, “seven V’s”, and even “ten V’s” - volume, variety, velocity, value, veracity, variability, validity, vulnerability, volatility, and visualization [28]. Also, in Bill Vorhies blog “How Many “V”s in Big Data – The Characteristics that Define Big Data” are pointed viscosity (mentioned as velocity element) and virality V’s [29]. The semantics of big data determined by problem domain of big data content was received from. Therefore, methods for big data processing must consider type, characteristics and semantics of big data. Semantics of big data is depended of primary disaster factor.

5.3. Big Data Maturity Model
Big Data Maturity Model (BDMM) help organizations to create information technology around their Big Data capabilities.
The goals of BDMMs are:
- To provide a capability assessment tool that generates specific focus on big data in key organizational areas
- To help guide development milestones
- To avoid pitfalls in establishing and building big data capabilities

Big data maturity models can be broken down into three broad categories namely:
- Descriptive
- Comparative
- Prescriptive models.

5.4. Big data platforms
Big data platforms are differed by:
- Processing approach;
- Resource management;
- Manner of data warehousing;
- Access methods.

5.5. Hadoop Spark and other frameworks
The technology of big data processing includes such stages as data generation and/or acquisition, big data storage, big data analysis and interpretation, big data application, big data archiving. Each stage of processing is supporting by appropriate platforms.

5.6. Big data can be processed by different platforms and different manners.
One of advanced platforms of big data processing is Hadoop (Welcome to Apache Hadoop [30]; Cloudera Distribution for Hadoop (CDH) [31]; Hortonworks Data Platform (HDP) [32]; MapR [33]. The base Apache Hadoop framework is composed of the following modules:
– Hadoop Common – contains libraries and utilities needed by other Hadoop modules;
– Hadoop Distributed File System (HDFS) – a distributed file-system that stores data on commodity machines, providing very high aggregate bandwidth across the cluster;
– Hadoop YARN – a resource-management platform responsible for managing computing resources in clusters and using them for scheduling of users' applications;

– Hadoop MapReduce – an implementation of the MapReduce programming model for large scale data processing.

5.7. Disasters and big data

Each type of disaster information has specific format and semantics. Data and information about disasters have the distributed and sporadic nature and arrive from variety of different sources. So this data is the big data.

Big data processing implies using of appropriate processing model. Most prevalent model for big data processing is map/reduce model. We use map/reduce notion to differ abstract programming model from its implementation MapReduce for processing big data.

Information about disasters came from different sources and contains variety data. Map/reduce model require defining nodes for map procedure (Fig.3).

5.8. Disaster risk awareness

Various big data platforms should be integrated in analytical systems of disasters awareness and risk reduction.

Using agents in big data processing for prediction, prevention, risk reduction, and mitigation of disasters. In consideration complexity of big data processing it’s proposed using of agents communities for managing big data. Intelligent agents classified into five classes [34] based on their degree of perceived intelligence and capability:

1) simple reflex agents;
2) model-based reflex agents;
3) goal-based agents;
4) utility-based agents;
5) learning agents.

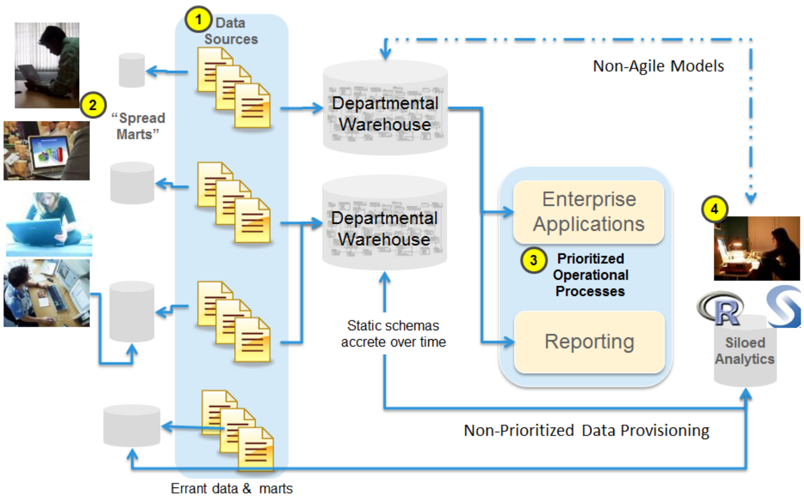

Figure 3. A big data aggregation in disasters risk management systems

Coordinated agent communities form multiagent system that may consists of four classes of agent architectures [34]:

1) Logic-based agents - in which the decision about what action to perform is made via logical deduction;
2) Reactive agents - in which decision making is implemented in some form of direct mapping from situation to action;
3) Belief-desire-intention agents - in which decision making depends upon the manipulation of data structures representing the beliefs, desires, and intentions of the agent;

4) Layered architectures - in which decision making is realized via various software layers, each of which is more or less explicitly reasoning about the environment at different levels of abstraction.

5.9. Technology of aggregation of big data

On the basis of the above the information technology of big data aggregation assumes appropriate structure. Big data processing consists of such stages: receiving data from the source – data recognition/sorting – awareness of data meaning – knowledge query – choosing data science method and big data algorithm (probably with machine learning) – choosing approach and platform of data processing – data processing – data visualization – data interpretation – decision making – disaster risk management (situation management). Each stage of processing considers transformation of data types of previous stage. Hence, the structure of big data technology should maintain all these stages (Fig.3).

6. Future Research - a combination of virtual reality, artificial intelligence and big datain

The future work will delve into the visualization of big data with immersive technologies, such as Virtual Reality (VR) and Augmented Reality (AR) (with possible continuation to Extended Reality, XR), and the application of Artificial Intelligence. The combination of Virtual Reality and Artificial Intelligence will offer a new approach to data analytics and thus will make it compelling for organizations to implement data-driven science in the disaster management training units.

7. Conclusion

The proposed approach to BigData aggregation for the disasters awareness is based on agents communities work. Such agents communities should be grouped according to appropriate layer services and may use knowledge based on ontology about own layer. In paper were determined layers of receiving and adoption information about disasters, classification and storing this information for further using. Thus, the awareness of the disaster situation is context-dependent and is determined by the knowledge and subjective assessment of the expert on the state of the environment for disaster management. To create agents' community based on knowledge about the situation and environment of situational management, it is necessary to formalize their description. Formalization of the composition of aspects of disaster management is carried out using ontologies. Accounting in the formal models of modality allows us to clarify the semantic context of the situation. Consolidating situation data on an ontology basis with modalities allows to create an adequate model of the situation and to form a system vision of the problem for the implementation of situational management. Proposed models may be used for constructing a network of disaster management situational centers with a well-grounded list of services for disaster management of a particular application.

8. References

[1] Disaster –http://www.preventionweb.net/english/professional/terminology/v.php?id=475
[2] Disaster risk management, www.preventionweb.net/english/professional/terminology
[3] Situation. Merriam-Webster Dictionary. http://www.merriam-webster.com/dictionary/situation
[4] Jakobson G, Buford J, Lewis L 2007 Situation Management: Basic Concepts and Approaches. In: Popovich, Schrenk, Korolenko (eds.) *Information Fusion and Geographic Information Systems* 2007. LNG&C, vol. XIV, pp 18-33, Heidelberg: Springer
[5] Naderpour M, Lu J, and Zhang G 2014 An intelligent situation awareness support system for safety-critical environments. *Decision Support Systems* **59** pp 325-340
[6] Disaster Management: Types, Awareness and Schemes for Disaster Management. http://www.yourarticlelibrary.com/disasters/disaster-management/disaster-management-types-awareness-and-schemes-for-disaster-management/30169/
[7] International Federation of Red Cross and Red Crescent Societies 2011 *Public awareness and public education for disaster risk reduction: a guide.* https://www.ifrc.org/Global/Publications/disasters/reducing_risks/302200-Public-awareness-DDR-guide-EN.pdf
[8] Techopedia 2020 *Data aggregation.* www.techopedia.com/definition/14647/data-aggregation
[9] Larose D 2006 *Data mining methods & models*. John Wiley & Sons.

[10] Assunção M, et al. 2015 Big Data computing and clouds: Trends and future directions. *J. of Parallel and Distributed Computing* **79** pp 3-15.
[11] Chen Y, et al. 2016 Big data analytics and big data science: a survey. *Journal of Management Analytics* **3** (1) pp 1-42
[12] Arbnor I, and Bjerke B 2008 *Methodology for creating business knowledge*. Sage Publ. Ltd.
[13] King W 2009 *Knowledge management and organizational learning*. Springer US, pp 3-13
[14] Kovalenko O 2014 Knowledge Models for Organizational Maintenance of Situation Centers. *Proceedings of 4th International Conference on Application of Information and Communication Technology and Statistics in Economy and Education (ICAICTSEE–2014)*, October 24–2, 2014, UNWE, Sofia, Bulgaria, pp 241-248
[15] Schutt R and O'Neil C 2014 *Doing data science: Straight talk from the frontline*. O'Reilly Media
[16] Dhar V 2013 Data science and prediction. *Communications of the ACM* **56** (12) pp 64-73
[17] Martin M E and Schuurman N 2017 Area-based topic modeling and visualization of social media for qualitative GIS. *Annals of the American Association of Geographers* **107**(5), pp 1028–1039
[18] Wooldridge M 2002 *An Introduction to Multi-Agent Systems*. John Wiley and Sons, NY.
[19] Buford J, Jakobson G, Lewis L 2006 Multi-Agent Situation Management for Supporting Large-Scale Disaster Relief Operations. *Int. J. of Intelligent Control and Systems* **11** (4) pp 284-295
[20] Kovalenko O 2014 Models and Means for Service Agents Orchestration in Situation Management Systems. *Actual Problems of Economics* **4** pp 462-467
[21] D'Aniello G, Loia V and Orciuoli F 2015 A multi-agent fuzzy consensus model in a situation awareness framework. *Applied Soft Computing* **30** pp 430-440
[22] Oxford Dictionary, knowing, knowledge, https://en.oxforddictionaries.com/definition
[23] Kempe Sh 2016 Data Modeling for Big Data, https://www.dataversity.net/ldm-webinar-data-modeling-big-data/
[24] DCMI Usage Board 2019 *Dublin Core Metadata Element Set, Reference Description*, http://dublincore.org/documents/dces
[25] UK Data Archive 2020, *Standards-tools*, https://www.data-archive.ac.uk/
[26] US EPA 2020 System of Registries (SoR), https://ofmpub.epa.gov/sor_internet/registry/sysofreg/
[27] PNNL 2008 Framework for Risk Analysis Multimedia Environmental Systems, https://mepas.pnnl.gov/FRAMESV1
[28] Firican G 2017 *The 10 Vs of Big Data*, https://tdwi.org/%20Articles/2017/02/08/10-Vs-of-Big-Data.aspx?Page=1
[29] Vorhies B 2013 *How Many "V"s in Big Data – The Characteristics that Define Big Data*, http://data-magnum.com/how-many-vs-in-big-data-the-characteristics-that-define-big-data/
[30] Apache Hadoop 2020 *Welcome to Apache Hadoop*, http://hadoop.apache.org
[31] Cloudera 2020 *Cloudera Distribution for Hadoop*, http://www.cloudera.com/ products/apache-hadoop/key-cdh-components.html
[32] Hortonworks 2020 *Hortonworks Data Platform*, http://hortonworks.com/products/data-center/hdp/
[33] MapR 2020 *Why-mapr*, https://www.mapr.com/why-hadoop/why-mapr
[34] Weiss G 2013 *Multiagent systems* (2nd ed.) Cambridge, MA: The MIT Press