BIM ORIENTED INTELLIGENT DATA MINING AND REPRESENTATION

Jiarui Lin, PhD Candidate, linjiarui.thu@gmail.com
Zhenzhong Hu, Associate Professor, huzhenzhong@tsinghua.edu.cn
Jianping Zhang, Professor, zhangjp@tsinghua.edu.cn
Department of Civil Engineering, Tsinghua University, Beijing, China

ABSTRACT

Building Information Model (BIM) encompasses mass of data specified by Industry Foundation Classes (IFC) from multiple fields during the whole lifecycle of the Architecture, Engineering and Construction (AEC) project. Recently, a large international initiative is launched to provide extensive support that will facilitate the creation, sharing and integration for BIM through all professions in the AEC industry, while few bend their mind to intelligent BIM data mining, which will improve the value of BIM. As the Information becoming increasingly rich and complicated, utilization of the data is getting harder, particularly for personnel without extensive knowledge of IFC specifications.

This paper proposes a framework utilizing Natural Language Processing (NLP) and International Framework for Dictionaries (IFD) to address the intelligent BIM data mining problem. First, user requirements depicted with a sentence of natural language are processed for keywords extraction by NLP. Then, keywords will map to IFC entities (or properties) through IFD-driven mapping mechanism, providing information for data retrieval and analyzing. Finally, the analyzed results of BIM data collected by previous research and application will be represented in accordance with its format such as tables, charts, animations or combination of them. The framework provides an intelligent data mining and multi-aspect data representation method for users without any special training, thus enhancing the applicable value of BIM.

Practical application results in construction management illustrates that with semantic understanding of his/her intention in natural language, user concerned data will be automatically retrieved, analyzed and represented in a suitable form, which is of great benefit for corporations without requiring extremely technological users, facilitating BIM application and enhancing the value of BIM.

Keywords: data mining, semantic analysis, Building Information Model (BIM), Natural Language Processing (NLP), International Framework for Dictionaries (IFD)

1. INTRODUCTION

Building information modeling (BIM) represents a new paradigm within architecture, engineering, and construction (AEC) industry, encouraging integration of the roles of all stakeholders in a project(Azhar 2011). According to reports from McGraw-Hill Construction and Pike Research, more and more engineers /designers /projects /companies are using BIM technology, as the year progresses, BIM adoption rate among AEC companies will continue to rise (Jennifer 2012).

As information of diverse disciplines constantly integrated into BIM model during the whole lifecycle of AEC project, BIM data becoming larger and larger. In today’s age of big data and multiparty collaboration there is an increasingly large amount of information created at any phase of AEC project, and companies need to deal with the deluge of information (Jasper 2011). It is difficult to manage, share and utilize the data of BIM. Industry Foundation Classes (IFC) is a comprehensive representation of the building model and the rules and protocols defining building data put forward by International Alliance of Interoperability (IAI, also known as buildingSMART), addressing the inadequate interoperability problem (Aranda-Mena and Wakefield 2006).

Much effort has been devoted to promoting the creation, sharing and integration of BIM as well as information and Knowledge Management (KM) throughout lifecycle of an AEC project, whereas few paid their attention to data mining and representation of BIM. Since AEC projects routinely involve multiple professionals from different functional areas and disciplines, effective and efficient communication of project information between diverse project participants is required. As concluded by (Anumba et al. 2008), explicit domain ontologies play an import role in supporting collaboration and decision making in various aspects of design and construction by generating a common representation of the problem-solving domain for the various disciplines. A theoretical framework of technical requirements for using BIM-server as a multi-disciplinary collaboration Platform was also discussed (Singh, Gu and Wang 2011). Procedures for developing Information Delivery Manual (IDM) were defined (Eastman et al. 2009), thus providing precise exchange of BIM data. A new approach to achieving interoperability between Web-based construction products catalogues was presented by (Kong et al. 2005), while (Redmond et al. 2012) discusses how information exchange can be enhanced by Cloud BIM. BIM applications including cost analysis of concept design (Abdelmohsen, Lee and Eastman 2011), construction simulation and safety analysis (Zhang and Hu 2011) were also explored. However, as the data size and functions of BIM software keep increasing, it is more difficult for users to get information they concerned in a proper way, things become worse even when they have less knowledge about IFC. With the increasing popularity of mobile terminals, simple and convenient human-computer interaction will be the future trend, meanwhile, research interests on cloud computing continue to rise. There is an urgent demand as well as a good foundation for intelligent BIM data mining and representation according to user's intention, which would help to simplify human-computer interaction.

In this paper, an framework for intelligent BIM data mining and representation was established. First, natural language processing (NLP) was proposed to process user's intention in natural language, from which keywords user concerned were extracted and then mapping to IFC Entities (or attributes), thus providing information for how to retrieve the related data in BIM database. Finally the data was analyzed and represented by charts, tables .etc. In section 2, correlational research was reviewed and the possibility of intelligent BIM data mining and representation was discussed. Detail processes of Keyword extraction based on NLP and its mapping mechanics to IFC entities supported by International Framework for Dictionaries (IFD) are described in section 3, while section 4 illustrates how information is retrieved, analyzed and represented. After case studying in section 5, the validity and practicality of the proposed approach are examined, conclusions in section 6 were obtained.

2. LITERATURE REVIEW

2.1 BIM, IFC, and IFD

As is clear from the definition (NBIMS 2012) from the National Building Information Modeling Standards Committee of USA that BIM encompasses information throughout the lifecycle of a facility, supporting multi-disciplinary collaboration as well as decision making. IFC is a data protocol proposed for sharing information between different software. With years of development and improvement, IFC standard is widely accepted for information exchange, some IFC-based BIM servers (Beetz et al. 2010; J O Rgensen et al. 2008; Kang and Lee 2009) were also developed for information sharing, extraction and integration. Research on extending IFC by eXtensible Markup Language (XML), and information exchanges with help of model view (Eastman et al. 2009; Fu et al. 2006) or IDM (Zhang et al. 2012; Kim et al. 2010) were also discussed.

Schemas such as IFC define the way in which different BIM software applications communicate with each other. However, as stated in IFD white paper (BuildingSMART 2008), a controlled vocabulary of construction terminology, which IFC is lack of, is essential to support data exchange. With support of terminology in multi-language, IFD also provides a mapping mechanics from concepts to IFC entities and attributes (Zhang et al. 2012), supporting distinguishing concepts from specific linguistic instances. (Shayeganfar et al. 2008) proposed a case study on how to implement an IFD library using semantic web technologies, bridging the gap between building information models and web services. Thus, with support of IFD library and semantic web, terminology or its synonym in specific language can easily map to the entity in a data schema like IFC.

2.2 Natural Language Processing

Natural Language Processing (NLP) is the computerized approach to analyzing text that is based on both a set of theories and a set of technologies (Liddy 2001), which is widely used in medicine, Artificial Intelligence (AI) and other areas. Broadly construed, NLP is considered to involve (CCSI 2012): (1) **Signal processing** takes spoken words as input and turns it into text. (2) **Tokenization or Word Segmentation** is the conversion of an input signal into parts (tokens) so that the computer can process it. (3) **Syntactic analysis** gets at the structure or grammar of the sentences. (4) **Semantic analysis** deals with the meaning of words and sentences, the ways that words and sentences refer to elements in the world. (5) **Pragmatics** concerns how the meaning of a sentence depends on its function in everyday life.

As an essential part of NLP, tokenization for different language was deeply discussed (Chiarcos, Ritz and Stede 2012; Wu and Fung 1994). Meanwhile, tools including Stanford parser (De Marneffe, MacCartney and Manning 2006), NLTK (Bird 2006) were developed for tokenization as well as other NLP needs. Since Chinese has no symbols (like spaces or punctuation in English) separating words, Stanford parser and ICTCLAS (Zhang et al. 2003) also provide components for Chinese language processing.

In addition to tokenization, the processing of NLP involves syntactic tagging and grammar analysis. Thus, relations between different segments in a sentence were determined. After processed by NLP, user's intention in a sentence is parsed to tokens dependent on each other, laid the foundation for their mapping to IFC entities.

2.3 Data mining

Data mining is the search for valuable information in large volumes of data (Weiss and Indurkhya 1998). The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Data mining is mainly composed of two parts: data pre-processing and data analysis. The former retrieves data from various data set and transforms data for data analysis, while the later analysis the data with different algorithms. Nowadays, data mining has been used for medical and spatial data analysis as well as for knowledge discovery.

Six common classes of tasks were involved in data mining as stated by (Fayyad, Piatetsky-Shapiro and Smyth 1996), they are Anomaly detection, Association, Clustering, Classification , Regression, and Summarization.

2.4 Discussion

Since there are already lots of research about NLP and a series of sophisticated tools have come forward, things become easy to process user intention and show the responding result based on plain text or speech as siri do. IFC, an object-oriented and semantic schema, provides a solid foundation for BIM data storage and interoperability. In addition, by years of practice, a large amount of BIM data were Accumulated and will keep increasing. Different storage and sharing platform for BIM were also developed and tested in several projects. Therefore, it is very meaningful to further promoting the value of BIM based on analyzing the big data of BIM. Furthermore, research results of data mining and visualization are of significance for BIM data mining and representation.

3. KEYWORDS EXTRACTION AND MAPPING

Keywords are what users care about when describing their intention and should map to entities or attributes in a data schema like IFC. Keywords extraction and mapping laid the foundation for user intention understanding and BIM data retrieving. In order to deal with keywords extraction and mapping, five steps were proposed as following (see Figure 1):

(1) Tokenization: segment user input into words, different NLP algorithms (such as Markov models) will be used when processing text in different languages;

(2) Tagging: labeling each word of the sentence as noun, verb or adjective, etc.

(3) Parsing: based on part of speech tagging, analysis syntactic structure of the sentence, thus getting the relationship between different segments of the sentence.

(4) Classification: according to its part of speech, each word will be sorted into different collection, which helps in determining whether a word should map to entity or its attribute, these words were called “keywords”.

(5) Mapping: each keyword is mapped to entity or its attribute in accordance with the relationships between concepts in IFD and entities in a data schema.

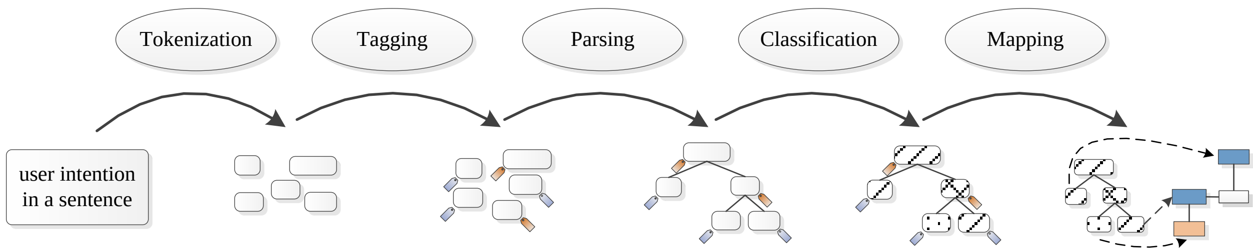

Figure 1: steps of keywords extraction and mapping

3.1 Keywords extraction based on NLP

By keywords extraction , the following questions should be cleared: (1) what kind of information is the user concerned? (2) what types of data should be retrieved from the database and (3) what about the constraints when retrieving data?

Take sentence “quantity of beams of second and third storey” as an example, following the steps described above, user’s intention in a sentence processed by NLP tools (NLTK, Stanford parser or other tools) will be a form similar to Figure 2: the sentence was split into words, and all words were tagged by Penn Treebank POS tag set (Bies et al. 1995) as well as the syntactic structure shown as a tree.

From root of the tree, we can find a path goes through all node tagged as NP/NN (S)/NNP (S) to the leaf (left blue part of Figure 2), which represents what the user wants finally and was called keyword. What’s more, for each sub-tree whose root is tagged as noun (or other noun phrase), the path to leaf with all nodes tagged as noun or its equivalent tag determines the keyword of this sub-tree, and the child tagged as adjective or preposition phrase performs as a constraint for the keyword (middle green part of Figure 2). Then, as shown in Figure 2, we can easily know that what the user wants is the quantities constraint to beams, which were contained in a storey whose name is equal to second or third. These grammatical relationships are of great value for the subsequent data retrieval and representation.

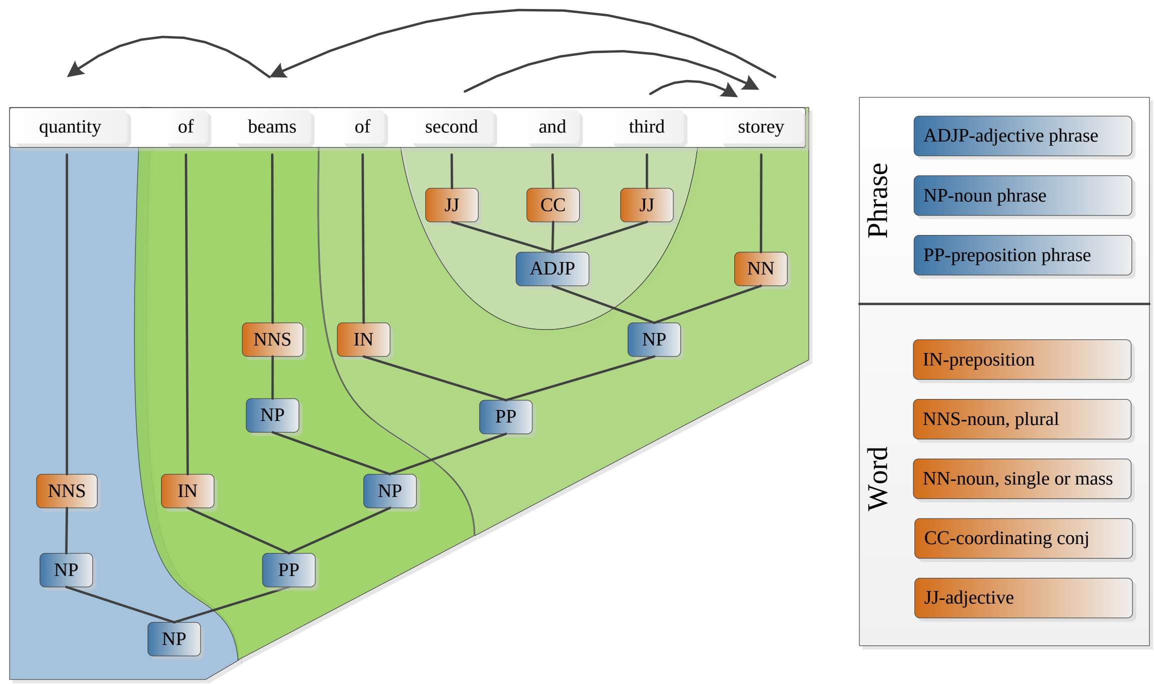

Figure 2: tags of segments and their relations in a sentence

3.2 IFD-based Keyword mapping

IFD Library is an open library, where concepts and terms are defined, semantically described and given a unique identification number. In order to make the information exchanged through IFC understandable, this information should rely on concepts as defined in IFD. Not only a synonym or plural form of a name of an entity but also entity's name in different language can be correctly understood, as long as the correct GUID is given. Based on analysis of quantity takeoff standard in China, the prototype of IFD developed before (Zhang et al. 2012) was extended in the following two aspects in order to support keyword mapping:

(1) synonym expansion: terms with the same meaning were related to the same concept by object relationships. Thus, the word "girder" and "beam" should all be related to the term "beam".

(2) form standardization: different forms of a term, including single/ plural form, different spelling, abbreviations were also related to a "standard" term.

On the basis of mapping mechanics provided by IFD, extracted keywords could be mapped to IFC entities or attributes. First, as keywords may be in different forms, all keywords are replaced by their "standard" form, and then, a synonym checking is needed. Finally, with the relationships between concepts and IFC entities and attributes, keywords can be mapped to IFC entities and attributes, thereby supporting data retrieval. Thus, the words "quantity", "beams", and "storey" will be mapped to "IfcProperty", "IfcBeam", and "IfcBuildingStorey" respectively. Note that due to limitations on space, discussion about how to map an abstract term like "quantity" (means weight when material is steel while volume when material is concrete) was omitted.

4. DATA MINING AND REPRESENTATION

4.1 Data retrieval and mining

As for data retrieval method based on IFC, a model-based system was developed for extracting structured information (Caldas and Soibelman 2003), BIM servers like IFC server also can retrieval BIM data from a database. However, for an IFC-based database, to achieve the purpose of data retrieval, relationships between entities in IFC schema are still needed. Since, method for keyword extraction and mapping proposed in section 3 has established the relationships between keywords and entities in IFC schema, a graph-based path search method was adopted to find the relationships between entities in a data schema.

An express file converter was developed, so that we can convert an IFC schema to a graph for entity relationship searching. When converting the IFC schema, all entities and defined types were taken as nodes while attributes and inheritances were considered as edges. At the same time, attributes of deriving from its super type were also copied to the entity for convenience. The express file for IFC schema was converted to a xgml file for graph storage as shown at the left of Figure 3.

Since all relationships were described in the graph, giving some entities, a path can be find based on path search algorithm like Dijkstra. Taken sentence in Figure 2 as an example, the path connecting “IfcBuildingStorey”, “IfcBeam”, and “IfcProperty” was found as the right of Figure 3.

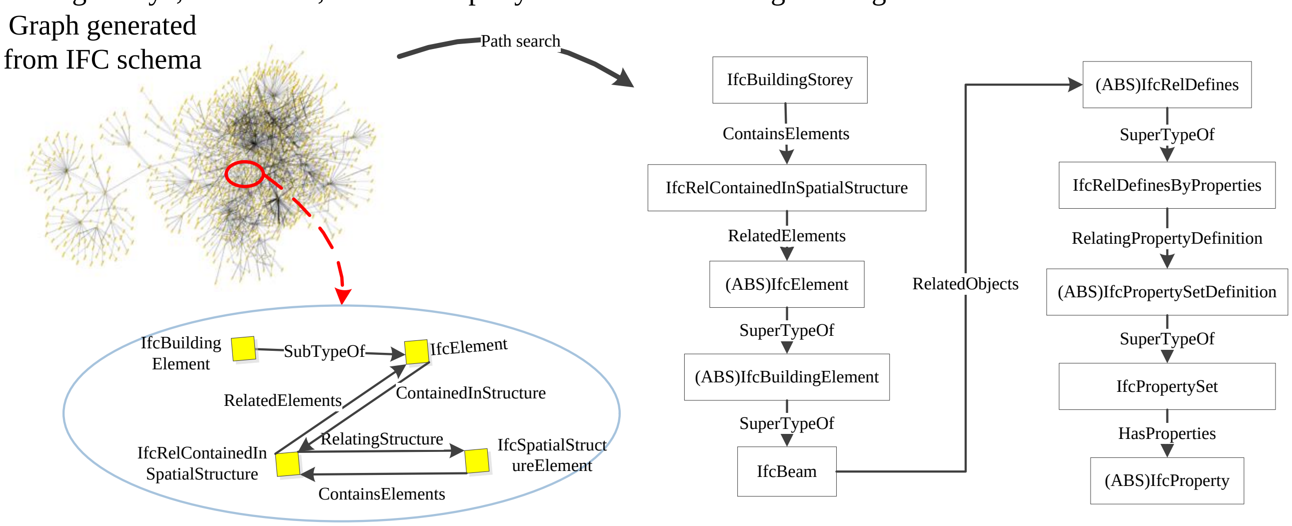

Figure 3: data retrieval path found in graph converted from an IFC schema

With relationships found, data retrieval scripts can be generated automatically according to different data storage strategies of different IFC-based BIM servers. In this paper, as the BIM server taken SQL Server as data storage, so SQL scripts were generated, then the data was extracted from the database.

Data retrieved from database was tagged with keywords extracted in section 3 as a key-value format, thereby the data can be grouped into different parts for data mining. In this paper, two main methods of data mining were adopted.

(1) Classification: as all data retrieved was tagged with keywords, all the data can be classified into different groups by keywords. As in the example mentioned above, quantities of beams with different materials first were classified by material into “weight” group and “volume” group; then, each of the groups can be classified by the storeys that beams contained in.

(2) Summarization: once the data retrieved has been classified into different groups, we can give an summary for each of the groups; for quantities in the same group we can easily sum the data.

4.2 Data representation

Data representation encompasses formatting data into tables, tree lists as well as visualizing the data as different kinds of charts, which has a significant impact on how user understand the data. Research on visualization of Categorical data (Chang and Ding 2005), time oriented data (Klimov, Shahar and Taieb-Maimon 2010) and so on were discussed. According to application and analysis on BIM data representation in the early researches, different data formats were summarized as follows:

(1) Single value: the simplest format, like height of a beam.

(2) Array: a collection of objects in the same format.

(3) Tree: a set of objects in same format, and part of them can be nested in one of the collection, tree format in IFC includes IfcSite, IfcBuilding and IfcSpace.

(4) Net structure: objects that related to each other, like IfcProcess in IFC.

(5) Composition of above

In addition, in order to help user to get a better form of the data, representation methods recommended were listed in Table 1.

Table 1: representations for different data formats

Data format	Representation	Detail
Single value	Plain text	Value can be highlight by coloring or bold font if value has an unit
Array	Charts & tables	1. One dimensional Array and two dimensional array contains string value can be visualized as column chart and bar chart 2. Other two dimensional array and higher dimensional array should be represented as chart with multi-series. 3. Time based data should take time as data for x axis. 4. Details of array can be listed in a table for reference.
Tree	Grouped charts & tree list	1. Data in a tree format may take chart in different details. 2. Tree list is a suitable way for details.
Net structure	Net graph	1. Best visualization for net data format is a net graph with labeling. 2. Schedule data should be visualized as gantt chart or timeline. 3. Details of the data can be in a table or tree list.
Data related to shapes	Charts & tables & 3D Geometries	1. Shapes which time based or data with sequence related to should be visualized as animation, while other geometries should be invisible. 2. Shapes other data related should be color according to different data related with other shapes invisible. 3. Data can be represented as mentioned above.

5. CASE STUDY

With method proposed above, an application called Intelli-BIM was developed. At the same time, practical testing was implemented on the data collected in the construction of the third stage of navy general hospital. With a total investment approximating 600 million US dollars, the hospital locates in the center of Beijing, China. Covering an area of over 100,000 m^2 and an aggregated floorage of 130,000 m^2, this project is comprised of 2 parts including a 15-storey 50m high medical building and an Integrated Research Building.

BIM model (including architecture, structure and MEP) in design phase was established by Revit 2011, then information about construction schedule and quantity takeoff was integrated by 4D-GCPSU (Hu and Zhang 2011), thus supporting 4D dynamic construction simulation, schedule management as well as resource management. From BIM data formed by 4D-GCPSU, the part of the medical building was picked out and ported into Intelli-BIM for testing. The Revit model and ported BIM model for user query in Intelli-BIM are shown in Figure 4. Total data size of BIM model for medical building is about 5 gigabytes including indices optimized for user query.

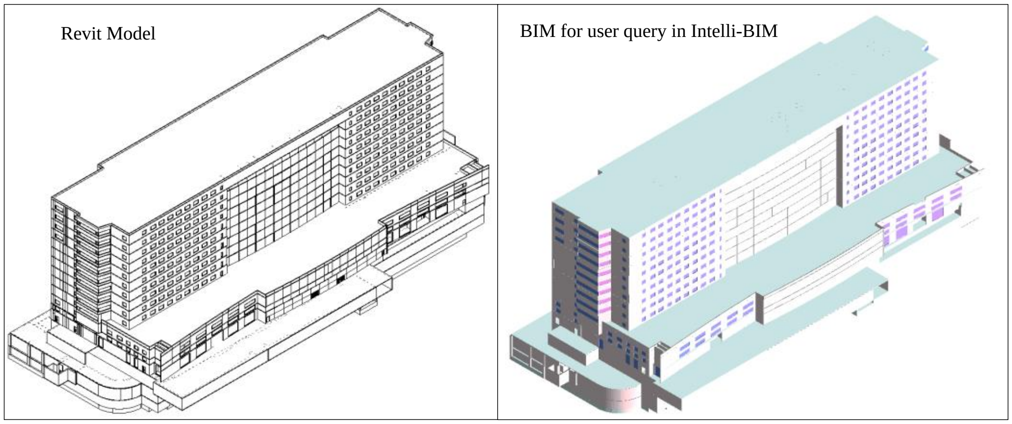

Figure 4: BIM model of medical building shown Revit and Intelli-BIM

Since all BIM data from design phase and collected in construction phase was ported into Intelli-BIM, once user intention described in a natural language sentence was input to the application, based on the method put forward above, user concerned data will be extracted, analyzed and represented in different forms. Take "quantity of beams of second and third storey" as an example, the sentence is processed by NLP for keyword extracting and keyword mapping through IFD, providing information for how to retrieve data from a BIM server, finally, the data will be represented in a diagram, a detailed list view and the related building elements will be filtered out as shown in Figure 5. Data and its related elements will be visualized in the same color according to the category of the elements (material of an element in this example) no matter when it is a diagram, a list view or a 3D view, thereby providing a consistent understanding for different representation forms.

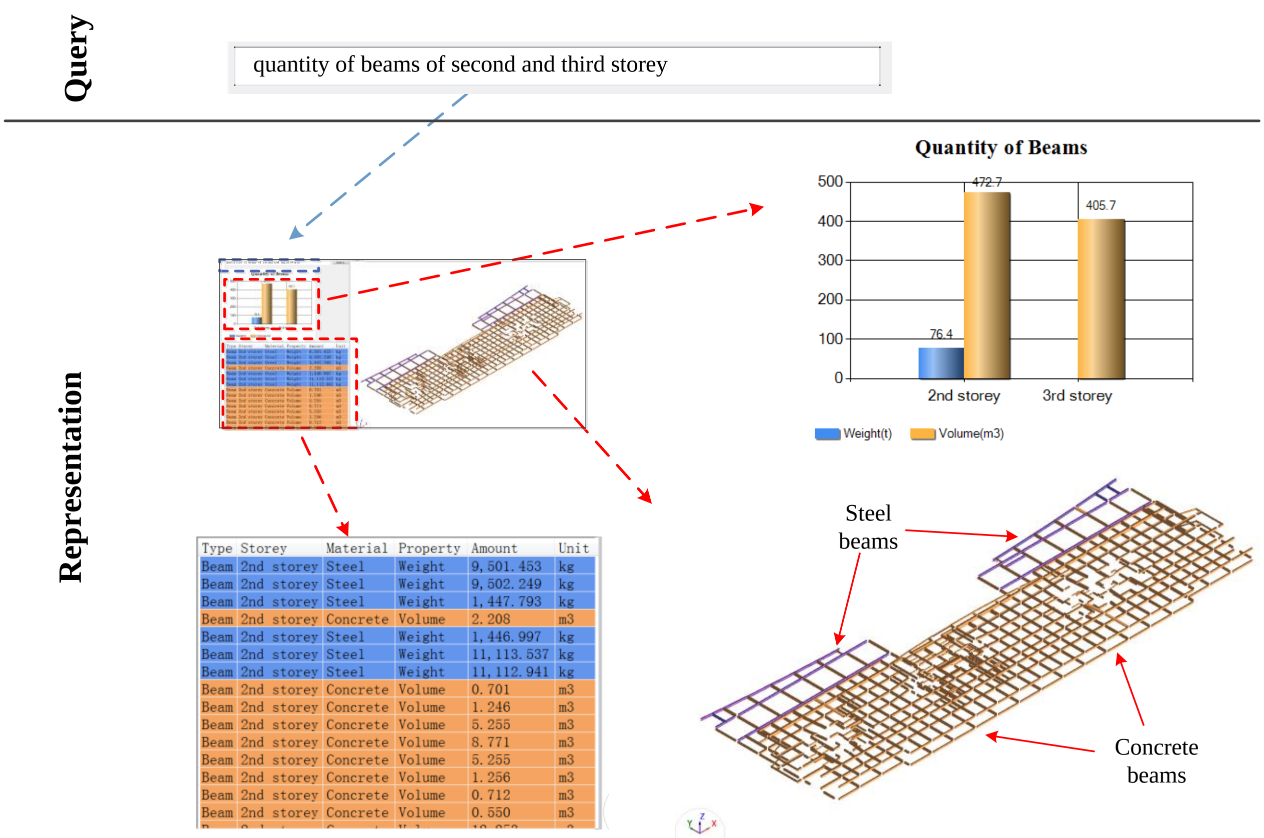

Figure 5: an example of BIM orient intelligent data mining and representation

6. CONCLUSION

In this paper, a novel framework for BIM oriented intelligent data mining and representation encompassing a set of enabling methods as well as a tool for intelligent and flexible BIM data mining and representation were introduced. First, an NLP-based method is incorporated in the keywords extraction from user's intention in natural language. Second, a mechanics is devised on the basis of IFD for mapping keywords to IFC entities or attributes. Thus, user concerned data could be retrieved from BIM database. Finally, the data will be represented according to its format such as tables, charts, animations or combination of them. Practical application results in construction management illustrates that with semantic understanding of his/her intention in natural language, user concerned data will be automatically retrieved, analyzed and represented in a suitable form, which is of great benefit for corporations without requiring extremely technological users, facilitating BIM application and enhancing the value of BIM.

The presented approach may be adopted in other applications such as facility management, cost management to simplify the human-computer interaction. However, lots of works still needed for enriching the content of IFD and its mapping mechanics to IFC schema as well as better data mining and representation methods.

ACKNOWLEDGMENTS

This paper was supported by the National Natural Science Foundation of China (No.51278274, No.51008168) and the Tsinghua University Initiative Scientific Research Program (No. 2011THZ03).

REFERENCES

Abdelmohsen, S. M., J. K. Lee, & Eastman, C. M. (2011). "Automated Cost Analysis of Concept Design BIM Models". *Designing Together: CAAD Futures, PUL*.

Anumba, C. J., Issa, R. RA., Pan, J., and Mutis, I. (2008). "Ontology-based information and knowledge management in construction". *Construction Innovation: Information, Process, Management* 8 (3):218-239.

Aranda-Mena, G., & Wakefield, R. (2006). "Interoperability of building information-Myth of reality".

Azhar, S. (2011). "Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry". *Leadership and Management in Engineering* 11 (3):241-252.

Beetz, J., van Berlo, L., de Laat, R., & van den Helm, P. (2010). "BIMserver. org–An open source IFC model server". *Proceedings of the CIP W78 conference*.

Bies, A., Ferguson, M., Katz, K., MacIntyre, R., Tredinnick, V., Kim, G., Marcinkiewicz, M. A., & Schasberger, B. (1995). "Bracketing guidelines for Treebank II style Penn Treebank project". *University of Pennsylvania*.

Bird, S. (2006). "NLTK: the natural language toolkit". *COLING-ACL '06, at Stroudsburg, PA, USA*.

BuildingSMART. "IFD Library White Paper". http:// www.ifd-library.org/images/IFD_Library_White_Paper_-2008-04-10_I_.pdf (5/10/2013, 2008).

Caldas, C., and Soibelman, L. (2003), "Integration of construction documents in IFC project models", *Proceedings of the Construction Research Congress*, American Society of Civil Engineers, Honolulu, HI.

CCSI. "Introduction to Natural Language Processing". http://www.mind.ilstu.edu/curriculum/protothinker/natural_language_processing.php (4/21/2013, 2012).

Chang, C. H., & Ding, Z. K. (2005). "Categorical data visualization and clustering using subjective factors". *Data \& Knowledge Engineering* 53 (3):243-262.

Chiarcos, C., Ritz, J., & Stede, M. (2012). "By all these lovely tokens... Merging conflicting tokenizations". *Language resources and evaluation* 46 (1):53-74.

De Marneffe, M. C., MacCartney, B., & Manning, C. D. (2006). "Generating typed dependency parses from phrase structure parses". *Proceedings of LREC*, 449-454.

Eastman, C. M., Jeong, Y. S., Sacks, R., & Kaner, I. (2009). "Exchange model and exchange object concepts for implementation of national BIM standards". *Journal of Computing in Civil Engineering* 24 (1):25-34.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). "From data mining to knowledge discovery in databases". *AI magazine* 17 (3):37.

Fu, C., Aouad, G., Lee, A., Mashall-Ponting, A., & Wu, Song. (2006). “IFC model viewer to support nD model application”. *Automation in Construction* 15 (2):178-185.
Hu, Z. Z., & Zhang, J. P. (2011). “BIM-and 4D-based integrated solution of analysis and management for conflicts and structural safety problems during construction: 2. Development and site trials”. *Automation in Construction* 20 (2):167-180.
Jørgensen, K. A., Skauge, J., Christiansson, P., Svidt, K., Sørensen, K. B., & Mitchell, J. (2008). “use of IFC model servers-modelling collaboration possibilities in practice”. *Department of Production, Aalborg University*.
Jasper, L. “BIM: Managing the data deluge”. http://www.aconex.com/blogs/2011/04/bim-managing-the-data-deluge-part-1.html (4/12/2013, 2011).
Jennifer, A. “Expected BIM Trends in 2013”, http://www10.aeccafe.com/blogs/aecsanjay/2012/12/20/expected-bim-trends-2013/ (4/12/2013, 2012).
Kang, H., and Lee, G. (2009). “Development of an object-relational ifc server”. *ICCEM/ICCPM*.
Kim, C., Kwon, S., You, S., & LIM, J. (2010). “Using Information Delivery Manual (IDM) for efficient exchange of building design information”. *27th International Symposium on Automation and Robotics in Construction (ISARC 2010)*.
Klimov, D., Shahar, Y., & Taieb-Maimon, M. (2010). “Intelligent visualization and exploration of time-oriented data of multiple patients”. *Artificial Intelligence in Medicine* 49 (1):11-31.
Kong, S. C. W., Li, H., Liang, Y., Hung, T., Anumba, C. J., and Chen, Z. (2005). “Web services enhanced interoperable construction products catalogue”, *Automation in Construction* 14 (3):343-352.
Liddy, E. D. (2001). “Natural language processing”, 2nd ed. *Encyclopedia of Library and Information Science, NY. Marcel Decker, Inc.*
NBIMS. “National Building Information Modeling Standard V2 - United States.”. http://www.nationalbimstandard.org/ (4/21/2013, 2012).
Redmond, A., Hore, A., Alshawi, M., & West, R. (2012). “Exploring how information exchanges can be enhanced through Cloud BIM”. *Automation in Construction* 24:175-183.
Shayeganfar, F., Mahdavi, A., Suter, G., Anjomshoaa, A., Zarli, A., & Scherer, R. (2008). “Implementation of an ifd library using semantic web technologies: A case study”. *Proc., ECPPM 2008 eWork and eBusiness in Architecture, Engineering and Construction*:539-544.
Singh, V., Gu, N., & Wang, X. Y. (2011). “A theoretical framework of a BIM-based multi-disciplinary collaboration platform”. *Automation in Construction* 20 (2):134-144.
Weiss, S. M., & Indurkhya, N. (1998). “Predictive data mining: a practical guide”. *Morgan Kaufmann.*
Wu, D., & Fung, P. (1994). “Improving Chinese tokenization with linguistic filters on statistical lexical acquisition”. *ANLC '94, at Stroudsburg, PA, USA*.
Zhang, H. P., Yu, H. K., Xiong, D. Y., & Liu, Q. (2003). “HHMM-based Chinese lexical analyzer ICTCLAS”. *Proceedings of the second SIGHAN workshop on Chinese language processing*, 184-187.
Zhang, J. P., & Hu, Z. Z. (2011). “BIM-and 4D-based integrated solution of analysis and management for conflicts and structural safety problems during construction: 1. Principles and methodologies”. *Automation in construction* 20 (2):155-166.
Zhang, J. P., Lin, J. R., Hu, Z. Z., & Yu, F. Q. (2012). “Research on IDM-based BIM process information exchange technology”. *14th International Conference on Computing in Civil and Building Engineering*, Moscow, Russia.