

# Automated generation of model view definitions using an context-aware approach MVD generation from civil engineering data standard

Tuyen Le <sup>1</sup>, H. David Jeong<sup>2</sup>

(To be submitted to ASCE Journal of Computing in Civil Engineering)

## ABSTRACT

Open data standards (e.g. LandXML, TransXML) have been widely recognized as a solution to the interoperability issue in exchanging digital data in the transportation sector. Since these schemas include rich sets of data types covering a wide range of disciplines across all project phases, model view definitions (MVDs) which define subsets of schema are required to specify what types of data to be shared in accordance with a specific exchange scenario. The traditional method for generating MVDs is time consuming and tedious as developers have to manually search for entities and attributes names that semantically match to the data exchange requirements. This paper presents a computational method that automatically map users' keywords to semantics-equivalent data labels (classes and attributes) in LandXML data schema. The study employs a lexical database of civil engineering terms to interpret users's intention from their keywords. The study also introduces a context-aware entity search algorithm that is able to find equivalent or most similar entities for a given keyword. The developed method has been experimented on a set of keywords extracted from an asset management manual. The experiment results show that the design algorithm is successful in generating partial LandXML branches from keywords.

---

<sup>1</sup>Ph.D. Candidate, Department of Civil, Construction and Environmental Engineering, Iowa State University. Ames, IA50011. E-mail: ttle@iastate.edu.

<sup>2</sup>Associate Professor, Department of Civil, Construction and Environmental Engineering, Iowa State University. Ames, IA 50011. E-mail: djeong@iastate.edu.

**Keywords:** Civil Project, Model View Definition, Civil Engineering Lexical Database, LandXML, Data Extraction, Context-Aware

## INTRODUCTION

Neutral data standards have been widely accepted as an effective solution to the interoperability issue in the construction industry. Several open standards have been proposed for various construction sectors, for instance Industry Foundation Classes (IFC) (buildingSMART 2015) and LandXML (landxml.org 2016). These standardized data models consist of rich sets of data elements covering various business processes and disciplines during the project life-cycle. Since a specific data exchange scenario needs only a subset of data, neutral data standards alone are insufficient to facilitate seamless digital data exchange among project stakeholders (Froese 2003; East et al. 2012). Digital datasets are large and complicated; thus, in order to querying data, the end user is required to have considerable programming skills and proper understanding of the source data schema. There is a need for formal definitions of schema subsets determining what types of data for specific data use cases.

The extraction of partial information models is a crucial requirement for the reuse of digital data generated through the life cycle of a infrastructure project. Digital infrastructure models consist of rich sets of data elements covering various business processes and disciplines. However, a specific data exchange scenario needs only a subset of data [(Froese 2003; East et al. 2012)]. Since digital datasets are presented only in computer-readable formats and mostly complicated; data extraction, especially from multiple sources, becomes a big burden on the end user [(?)]. Users are required to have considerable programming skills and properly understand the structure and meanings of each entity or attribute included in the source data schema. This is especially challenging for the civil domain since data are commonly gathered from multiple sources with potential conflicts of data format and terminology (entity and attribute names). Data driven decision making based on a wrong extracted dataset would likely lead to a wrong decision. Thus, there is a demand for compu-

tational techniques that can allow for automated extraction of data with minimized human interfere.

Model View Definition (MVD), a concept introduced by buildingSMART, aims to fulfill the above need. MVDs are subsets of a standard data schema for particular data use scenarios. The availability of these model views will underpin the extraction of data from complicated sets of data generated from the AEC industry. The civil sector has introduced several model views of the Landxml schema. The pioneer effort in this area is the InfraModel project carried out by the Technical Research Center of Finland that aims to specify subsets of LandXML schema for different types of transportation assets (inframodel.fi 2016). In spite of a considerable number of research have been made, the existing MVDs are still limited to a large demand from the industry. This is because that the current method for developing MVDs is on a manual basis which is time consuming (Venugopal et al. 2012b; Eastman 2012; Hu 2014). MVDs are required to be regularly tailored to reflect changes in industry practices. Therefore, it is demanded to shift the current ad-hoc practice of model view definition to a more rigorous methodology (Venugopal et al. 2012b).

To fill that need, this research aims to develop a computational framework for automated generation of MVDs for the civil sector. The platform is capable of interpreting users' intention from their input keywords and automatically generating corresponding subsets of LandXML schema. Keywords is the most preferred method for searching information and the preferred input they user like to use. As their effort in typing the input is minimized. This study proposed a method for generating mvd using user keywords. for large source schema like IFC which consist of thousands of entitieis and attributes, manually finding an equivalent items is tedious, time consuming and especially error-prone. in addition, a keyword meaning may be generic, it may contains detailed concepts. For example, if drainage is one of the needed information specified in an idm. The problem become extremely challenging when the current source schema does not contain the needed information, or their names are different from those mentioned in the idm. Specifically, the proposed framework is consisted of the

following two modules: (1) interpreting user’s data needs from their input keywords using a digital dictionary of transportation asset data elements, and (2) designing an context-aware algorithm that can find semantically equivalent or most similar data entities in the standard data schema. The advantage of this method is the understanding of senses of technical terms and match classes based on their semantics instead of their pure labels. It allows users to be flexible in keyword use and is expected to significantly reduce efforts on MVD development.

## BACKGROUND

### Partial model extraction

Methods for extracting partial models for specific use cases can be classified into the following groups ordered by the degree of ease of use for end users: (1) developing a query language specifically for Building Information Modeling (BIM) models, (2) ontology-based query approaches, and (3) user-oriented query methods. The first group aims to tailor the conventional query languages (e.g., SQL, Object Orientation) for extracting information from BIM models. The major focus is on developing spatial filter strategies. Examples of these efforts include the Spatial query language (?), QL4BIM Spatio-semantic query language (?); graph-based BIM retrieval (?), and topological querying (?). The second group is to enhance the human-readability of data schema by utilizing an ontology approach to transform relations among data entities from implicit to explicit. With these semantic presentations, it is easier for end users to read and comprehend a complicated data schema. An extensive number of studies based on this approach have been carried out for various use cases including ontology-driven construction information retrieval for tunnel projects (?), ontology partial BIM model extraction for building projects (?), ontology-based extraction of construction information (?) and ontology based querying over linked life cycle data spaces (?). The last class of partial model query approaches moves a step further in terms of enhancing the ease of data extraction by providing query tools that require less effort from users. For example, Won et al. (2013) (Won et al. 2013) proposed a no-schema algorithm that allows for the extraction of IFC instances without using IFC schema or MVD. In addition, a visual BIM

query (?) was also established to visualize query codes. Although significant research efforts have been conducted, there is still a lack of natural language interface platforms that can enable computers to understand and interpret the end user’s data interests in the civil infrastructure domain.

## **Data Standards for Civil Infrastructure**

A large body of research has been undertaken for the last two decades to establish open data standards for the highway industry. Most of the existing standards were developed adopting the XML technique. LandXML (landxml.org 2016), a result of early international collaboration efforts in facilitating interoperability in the civil industry, covers the following main groups of data: survey data, ground model, parcel map, alignment, roadway and pipe network. As an attempt to improve LandXML and propose a new standard specialized for the transportation industry, TransXML (NCHRP Project 20-64) project was chartered by the US National Cooperative Highway Research Program. TransXML focused on 4 business areas: survey/road design, construction/materials, bridge structures, and transportation safety (Scarponcini 2006). Of these domains, survey and geometric roadway classes are mainly derived from LandXML and are included suggestions for improvement (Ziering et al. 2007). buildSMART is also actively participating in developing standards for infrastructure assets. This agency has just released IFC 4x1 for alignment information and is carrying out several other ongoing projects such as IFC for Road and IFC for bridge.

## **Model View Definition**

Model View Definition (MVD) is a formal a subset of a data schema in accordance with data requirements for a specific data use case (See et al. 2012). MVD, which was first introduced by the building sector, is to ensure information to be used unambiguously when data is transferred from digital models to a domain user (Jiang et al. 2015). The concept of MVD has also been adopted by the infrastructure sector recently. The VTT Technical Research Center of Finland has proposed the Inframodel (inframodel.fi 2016) which is a Finish national application specification for subsets of LandXML schema. The current

version Inframodel 3 provides various MVDs for different types of infrastructure assets during the plan and design stages, for instance waterways, water supply and sewage, roadways and streets, railways, and pipe networks.

The process of developing a MVD includes the following three steps (See et al. 2012; Venugopal et al. 2012b; Eastman et al. 2009): (1) professional experts investigate industry business workflows and data exchange requirements to develop an Information Delivery Manual (IDM); (2) software developers translate the IDM in natural language into a computer-readable MVD by mapping required data to the entities in an open data schema and re-structuring them in a formal computerized format so that software vendors are able to develop data exchange applications; and (3) software applications are implemented and the translation results are validated by reviewing concept by concept. This approach to developing MVDs is on a manual basis which is time consuming (Venugopal et al. 2012b; Eastman 2012; Hu 2014). This leads to the shortage of MVDs in comparison with the large demand from the construction industry. Thus, there is a need for computational methods for automated generation of data schema subsets (Venugopal et al. 2012b).

## **Related Studies on Automated Generation of MVD**

A number of studies have been carried out to enhance the efficiency of developing MVD. Several researchers aimed to automatize step 2 by developing tools that allow for reduction of manual work by technical developers in translating IDMs into machine-readable MVDs, for instance the work by Katranuschkov et al. (2010). As an extension of this work into step 3, Windisch et al. (2012) proposed a framework that supports automated validation of model views. These tools can reduce burden on software developers; they, however, still are required to have a deep understanding of IFC schema and domain knowledge. Another research area is to improve the re-usability of MVD by introducing the concept of SEM (Semantic Exchange Model) (Venugopal et al. 2012a; Lee et al. 2016b; Zhang et al. 2013; Lee et al. 2016b). This approach develops modularized MVDs in which concepts and relations are explicitly structured using ontology. Recently, Lee et al. (2016a) proposed an ontology-based MVD

generator. The required data items are extracted from domain guidances and transformed into an ontology which is then mapped to the entities of IFC schema. By representing IDM in ontology format, this framework can reduce semantic mismatches between required data and IFC classes. However, these efforts are not yet fundamentally transform the way how MVDs are generated and maintained as much manual work is still needed for step 1 of the MVD development process. Generating MVDs using natural language input would provide ease of use to end users (Jiang et al. 2015); as a result, industry specialists can fully benefit from data-rich BIM models.

(Lee et al. 2016a): The suggested approach using a case study consists of four big steps: (1) establishing IDM ontology on the Protégé (ontology format of an written idm), (2) validating a dataset using semantic reasoning, (3) parsing OWL for generating concept modules and translating OWL/XML ontological IDM into mvdXML (using a map table mapping owl/xml and ifc entities), and (4) importing mvdXML into the IfcDoc tool to generate MVD documentation (ifcdoc can read mvdxml and generate mvd html docs).

## **Domain semantic concept mapping algorithms**

In the construction industry, research efforts are currently focusing on standardizing the data structure format, there are few research have been done to deal with the issue of sense ambiguity. Zhang and El-Gohary (2016) (Zhang and El-Gohary 2016) proposed an algorithm called ZESem aiming to match a certain keyword to the most semantic nearest IFC entity (extracted new concepts (then verified by users) and match them (those accepted by users- this is why it is called semiautomatic-adopted rate is around 80%) to existing ifc concepts (related if concept candidate are obtained using string-based and wordnet) through one of the following relations hypernym, hyponym and synonym). The algorithm includes two sequential steps including term-based matching and semantic relation based matching. Since the algorithm accepts matches from the label-based matching step, disambiguation still remains in cases in which the same word form is used for different senses. In addition, ZESem relies on Wordnet which is a generic lexicon, the applicability would be limited. Lin et al.

(2015) (Lin et al. 2015) developed a IFD based framework for BIM information retrieval. IFD Library (International Framework for Dictionaries library), which is developed and maintained by the international buildingSMART, is a dictionary of BIM data terminology that assigns synonyms the same ID. The integration or exchange of data using IDs rather than data names would eliminate semantic mismatch. However, since IFD is a hand-made electronic vocabulary, constructing this e-dictionary is time consuming and therefore it is still very limited to large collection of terms in the construction industry.

## KEYWORD-DRIVEN METHOD FOR GENERATION OF MVDS

Figure 1 shows the concept of the developed method for automated generation of MVDs. Using this approach, a subset of data schema will be generated for a given set of keywords. Those keywords describe data needs of specific contexts for instance bridge elements, drainage systems, land boundaries, traffic safety, alignment, pavement structure, etc. As shown in the literature review, LandXML-2.0 is the most comprehensive data standard for information exchange during the life-cycle of civil infrastructure assets. LandXML organizes classes and attributes in a tree structure. The purpose of this study is to propose a method that can assist users in extracting partial branches of entities from this neutral data standard using only keywords as input information.

The proposed algorithm for finding equivalent or most similar entities in LandXML schema is illustrated in Figure 2. The algorithm interprets users' intention from their keywords and look for semantics rather than specific words. As shown, the proposed matching algorithm is comprised of three stages. First, the algorithm generates a potential concept from the input keyword using a domain dictionary. In the second stage, the similarity between the input concept and every LandXML class is determined. The algorithm measures the similarity (S) based on two assessments including concept name match ( $S_n$ ) and context match ( $S_c$ ) given in Equation 1, where  $w_n$  and  $w_s$  respectively represent the weight of each matching type. Concept name matching is based on the similarity of the corresponding character sequences while context matching corresponds to how many attributes they share in



common. Finally, all the matched LandXML entities are ranked by the total similarity score, and the top match is recognized as an equivalent or the most similar entity to the user's keyword. The parent and children nodes connected to the selected one in the LandXML tree are detected and used for composing a partial schema. Sections below discuss in detail the process of generating concept candidate from a keyword and similarity measures.

$$S = w_n S_n + w_c S_c \quad (1)$$

## Domain knowledge base

in a previous study, the author develop a method for automated construction of domain knowledge, this study employed the open source toolkit previously developed by the authors. this method to develop a knowledge based for the civil engineering using civil engineering input corpus and the input. the process includes. the method is a machine learning based method which learning lexical relationship between terms based on their statistical data of their occurrence and their context words in domain text. the method enable to classify heterogeneous terms in the following relations: synonymy, hyponymy, and meronymy. Figure below shows a partial knowledge based network utilized in this study.

## Keyword to Concept Interpretation

Keywords are a short way that end users use to express their needs when searching for information. This study aims to search for semantically equivalent LandXML entities rather than specific words. Thus, interpreting semantics of keywords is a prerequisite task. For instance, using 'roadway alignment', the user intends to look for the specific 'alignment' attribute of the 'roadway' concept. In this study, a keyword is defined to be composed of the following two constituents: concept word (Cn), and context word (Cc). To detect these two elements from a keyword, the designed algorithm employs a civil engineering knowledge base namely Civil Engineering Lexicon (CeLex) developed by the authors (Le and Jeong 2016). This lexical database organizes civil engineering technical terms in a lexical structure where

terms are connected to each other through following relations: synonyms, attributes and hyponyms. By matching the individual words constituting a multi-word keyword to term names in CeLex, Cn and Cc can be determined. If the input is a single-word keyword, Cc is associated with a set of all CeLex terms that link to the broad Cn word via the attribute relation. As a result, given a keyword, this stage will generate a concept name Cn and its context words Cc(s).

### **Related concept collection**

this study employs different dictionary-based similarity to detect the most related keywords for a certain keyword. the related keywords could be belong to one of the following three categories, hyponymy (is-a) and meronymy (par-of).

### **Concept name matching - Sn**

Concept name is an indicator of semantic similarity. The degree of overlapping in name between the input concept and a source entity reflects the similarity degree between them. Tversky (1977) proposed that the similarity between two features can be a function of their common and distinguishing portions. Comparing two concept names can be based on the similarity of their characters. In this study, the name similarity are determined using Equation 2, where  $n_A$  and  $n_B$  are respectively the individual words of the target and source concept names. For example, the similarity between the ‘traffic sign’ keyword and the ‘road sign’ entity in LandXML is 1/3 (33%). To address the issue whereby there are variant names for representing the same concept, a set of synonym names extracted from the CeLex database for the input concept are used to find the most similar LandXML class. The final similarity score (Sc) is from the synonym that has the best match.

$$S_n = \frac{|Cn_A \cap Cn_B|}{|Cn_A \cup Cn_B|} \quad (2)$$

## Context matching - $S_c$

This measure compares context items of the target and source concepts. This is to reduce mismatch due to the issue of the same word referring to different things. By comparing their attributes and other related entities, their meaning difference can be detected. The similarity from this view can be measured by the commonality and difference between their context entities. In this study, the context of a source LandXML class includes its attributes and children entities in the LandXML tree hierarchy. The context of an input concept candidate, as discussed earlier, is composed of its attributes found in CeLex. The context similarity (as shown in Equation 3) is defined as the ratio between the number of common items and the total context words of the target concept, where  $C_{c_A}$ , and  $C_{c_B}$  respectively denotes the contexts of the input concept and a LandXML class.

$$S_c = \frac{|C_{c_A} \cap C_{c_B}|}{|C_{c_A}|} \quad (3)$$

## Distinguishing attribute measure

common attributes do not provide distinguishing difference for a concept. for example, with, length, area, type. These data attributes are common. Some distinguishing attributes of pavement include distress. In order to take into account this matter. The authors proposed the below method to measure the weight for each attribute. this measure is called t-value which is based on the frequency of each attribute in the lexicon.

$$t - value_k = 1 - \frac{|a_k|}{\max_{1 \leq j \leq n} |a_j|} \quad (4)$$

## IMPLEMENTATION AND DISCUSSION

## Experiment setup

to validate the system, 10 keywords representing general concepts in the civil engineering were selected randomly from subsection title in manual documents, since keyword in section title are likely to represent a certain interest context. Three Ph.D. students were invited to work as annotators who are asked to manually identify all the relevant matching entities in the which are supposed to be included in the gold standard MVD. A sub set of 100 entities in the schema were randomly selected. annotator are required to assign one of the two tags ‘related’ and ‘non-related’ to each pair of the input keyword and the the selected landxml schema.

the system was evaluated using evaluation measure adopted from the information retrieval domain. analogy, the number of entities retrieved is corresponding with the number of documents retrieved. the measure are defined using the following equations. to consider the ranking of th obtained result, reverse ranking are implemented.

An evaluation experiment was conducted to evaluate the context-aware searching algorithm. In this test, an asset management guidance was reviewed and eight keywords showing data requirements were extracted including roadway type, drainage, alignment, shoulder, pavement, safety, traffic lane, and climate. A graduate student was invited to manually identify branches of the LandXML schema that match to each of the tested keywords. The automatically generated MVDs were compared with the corresponding manually identified subsets to evaluate the algorithm performance.

## Results and discussions

Table 1 shows the experiment results when the weights  $w_n$  and  $w_c$  were both set 0.5. As shown in the table, the proposed method achieves a precision of 62.5% (ratio between correct answers and total tested input keywords). The algorithm successfully detects branches of LandXML tree even keywords are not similar to the names of the source entities. For example, ‘drainage system’ term is not used by LandXML, however the designed system can detect that the LandXML branch going through pipeNetwork entity is regarded to drainage

information. Since the algorithm performance would depend upon the size of sample keyword set and the selection of partial similarity weights, the authors intend to test the algorithm on a bigger dataset with various settings. One limitation of this research is that the algorithm generates only one relevant branch even for a broad keyword that can be associated to multiple nodes. Future research should be able to detect and aggregate multiple matched LandXML nodes for a single input.

## RESEARCH CONTRIBUTIONS AND IMPLICATIONS

the system developed in this study is expected to offer an effective supporting tools for mvd developers. the automated matching and related source entities generated by the system provide developers a short list of terms rather manually scrolling and investigating the meaning of each entities in the reference conceptual documents. By provide a list of the most semantically related items, focus is paid on those items and the time and efforts can be significantly reduced.

moreover, the system is expected to allow reduce time in idm development, in current practices, figure show the proposed mvd development using the proposed system. as shown, using this system, this approach is flexible, since the domain knowledge terminology has been captured and stored in the databases in terms of concepts relations, attributes. less effort is needed by the idm developer team in interviewing domain experts, documentsation and translating meeting minutes into spreadsheet. As shown the information format changing from spreadsheet to mvdxml. the translation is currently done by software developers. in this idm data exchange requirement documents. structuring information

## CONCLUSIONS

Model view definition has been widely recognized as a means for facilitating seamless information exchange throughout the project life-cycle. Although a large body of MVDs have been developed, they are still limited compared to the large demand in the construction industry. This is because that the current ad-hoc practices of MVD development is

time consuming. The contribution of this study is an automated method for generation of MVD using users' keywords. The designed algorithm leverages a domain data dictionary to interpret user's intentions. It also utilizes a context-aware approach to match the interpreted concepts to those entities in the LandXML data standard. The algorithm takes into account the variation in names of concepts, thus it can reduce mismatched due to the inconsistent use of terminology between users and the data standard. The proposed method was tested on a set of eight keywords extracted from an asset management manual. The result shows that the algorithm is an effective tool for extracting subsets of data schema. As using keywords is one preferable method for information search, the algorithm is expected to become a fundamental tool assisting professionals in extracting data from complex digital datasets.

## ACKNOWLEDGEMENT

This research was funded by the National Science Foundation (NSF) under the Award Number 1635309. The authors gratefully acknowledge NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

## REFERENCES

- buildingSMART (2015). "Ifc overview summary, <<http://www.buildingsmart-tech.org/>>." Accessed: 2015-10-11.
- East, E. W., Nisbet, N., and Liebich, T. (2012). "Facility management handover model view." *Journal of computing in civil engineering*, 27(1), 61–67.
- Eastman, C. (2012). "The future of ifc: Rationale and design of a sem ifc layer. Presentaion at the IDDS workshop.
- Eastman, C., Jeong, Y., Sacks, R., and 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.
- Froese, T. (2003). "Future directions for ifc-based interoperability." *ITcon*, 8, 231–246.

Hu, H. (2014). “Development of interoperable data protocol for integrated bridge project delivery.” Ph.d., Ph.d. UMI Dissertations.

inframodel.fi (2016). “Inframodel, <<http://www.inframodel.fi/en/>>. Accessed: 2016-12-10.

Jiang, Y., Yu, N., Ming, J., Lee, S., DeGraw, J., Yen, J., Messner, J., and Wu, D. (2015). “Automatic building information model query generation.” *Journal of Information Technology in Construction*.

Katranuschkov, P., Weise, M., Windisch, R., Fuchs, S., and Scherer, R. J. (2010). “Bim-based generation of multi-model views.” *CIB W78*.

landxml.org (2016). “About landxml.org, <<http://www.landxml.org/About.aspx>>. Accessed: 2015-12-10.

Le, T. and Jeong, H. (2016). “Technical term similarity model for natural language based data retrieval in civil infrastructure projects.” *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, Vol. 33.

Lee, Y.-C., Eastman, C. M., and Solihin, W. (2016a). “An ontology-based approach for developing data exchange requirements and model views of building information modeling.” *Advanced Engineering Informatics*, 30(3), 354–367.

Lee, Y. C., Eastman, C. M., Solihin, W., and See, R. (2016b). “Modularized rule-based validation of a bim model pertaining to model views.” *Automation in Construction*, 63, 1–11.

Lin, J.-R., Hu, Z.-Z., Zhang, J.-P., and Yu, F.-Q. (2015). “A natural-language-based approach to intelligent data retrieval and representation for cloud bim.” *Computer-Aided Civil and Infrastructure Engineering*.

Scarponcini, P. (2006). “Transxml: Establishing standards for transportation data exchange.” *Joint International Conference on Computing and Decision Making in Civil and Building Engineering*, Montreal, Canada (June).

See, R., Karlshoej, J., and Davis, D. (2012). “An integrated process for delivering ifc based data exchange.

- Tversky, A. (1977). "Features of similarity.." *Psychological review*, 84(4), 327.
- Venugopal, M., Eastman, C., and Sacks, R. (2012a). "Configurable model exchanges for the precast/pre-stressed concrete industry using semantic exchange modules (sem)." *International Conference on Computing in Civil Engineering*, 269–276.
- Venugopal, M., Eastman, C. M., Sacks, R., and Teizer, J. (2012b). "Semantics of model views for information exchanges using the industry foundation class schema." *Advanced Engineering Informatics*, 26(2), 411–428.
- Windisch, R., Katranuschkov, P., and Scherer, R. J. (2012). "A generic filter framework for consistent generation of bim-based model views." *European Group for Intelligent Computing in Engineering, EG-ICE 2012 - International Workshop: Intelligent Computing in Engineering*.
- Won, J., Lee, G., and Cho, C. (2013). "No-schema algorithm for extracting a partial model from an ifc instance model." *Journal of Computing in Civil Engineering*, 27(6), 585–592.
- Zhang, C., Beetz, J., and De Vries, B. (2013). "Towards model view definition on semantic level: A state of the art review." *European Group for Intelligent Computing in Engineering, EG-ICE 2013 - 20th International Workshop: Intelligent Computing in Engineering*.
- Zhang, J. and El-Gohary, N. (2016). "Extending building information models semiautomatically using semantic natural language processing techniques." *Journal of Computing in Civil Engineering*, C4016004.
- Ziering, E., Harrison, F., and Scarponcini, P. (2007). *TransXML: XML schemas for exchange of transportation data*, Vol. 576. Transportation Research Board.



TABLE 1: Experiment results

<b>Keyword</b>	<b>Matched Landxml Branches</b>	<b>S</b>	<b>Correct?</b>
drainage	PipeNetwork->pipeNetworkType->storm	0.25	yes
vertical alignment	Roadway->Alignment->Profile->ProfAlign	0.52	yes
shoulder	Roadway->Lanes	0.28	no
pavement	Zone->pavementSurfaceType	0.25	yes
roadway type	Roadways->Roadway->intersectionType	0.54	no
safety	Roadway->TrafficVolume	0.18	yes
	Roadway->Intersection->TrafficControl	0.18	yes
traffic lane	Roadway->Lanes	0.53	yes
climate	zoneMaterialType->soil	0.25	no

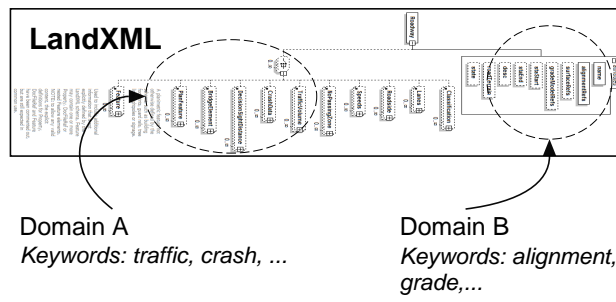


FIG. 1: Excerpts of Partial Views of LandXML

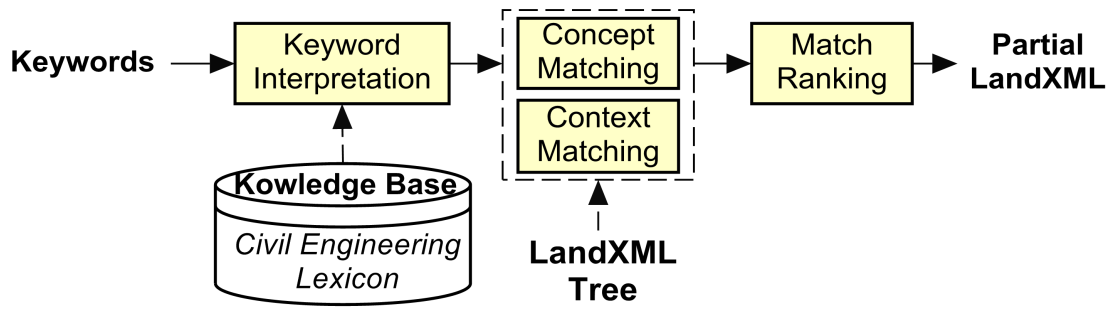


FIG. 2: Keyword-driven MVD generation approach