Rich Information Structures

TUNÇER Bige; STOUFFS Rudi; SARIYILDIZ Sevil
Faculty of Architecture, Delft University of Technology, The Netherlands
http://www.bk.tudelft.nl/informatica/toi/  b.tuncer@bk.tudelft.nl

Technological advances enable and encourage practitioners and students to make the design process more information intensive. This information intensity raises questions of complexity: how to organize and intra-relate large amounts of information in order to facilitate efficient retrieval of this information. This involves issues of both modeling and visualizing this complexity in design presentations and project documentation facilities.

We propose a methodology for constructing a rich information structure which offers new possibilities for accessing, viewing, and interpreting this information. Hereeto, we present two techniques: a decomposition of documents by content, and the separation of syntax and semantics. We then discuss the effects of both techniques on issues of flexibility, extensibility, and ease of use in constructing a rich information structure. We finally describe an exemplary application we are developing that combines the proposed methodology and techniques for the purpose of presenting architectural analyses.

**Keywords:** information structure; information modeling; extensibility; flexibility; ease of use.

Introduction

Technological advances enable and encourage practitioners and students to make the design process more information intensive and to exchange this information effectively with other participants in the process. This information intensity raises questions of complexity: how to organize large amounts of information and how to relate the information entities within the organization, in order to facilitate efficient retrievals of this information. This involves issues of both modeling and visualizing this complexity, in design presentations and project documentation facilities. Within web-based environments, a document-based approach is most commonly adopted, that is, information entities are treated as documents with associated attributes.

In order to discuss efficient and effective ways of retrieving information from such information structures, we must distinguish different levels of information retrieval. At one level, the purpose is to find and retrieve specific information entities. At another level, effective overviews of the information structure, or part thereof, enable interpretations of the information space that may lead to new understandings and to the recognition of important aspects or entities. Modeling the information structure with support for these levels and for various ways of information retrieval requires the construction of a rich information structure, that is, a tight structure of documents and relationships that enables a contextual interpretation beyond the information as contained in the individual entities. Such a rich structure offers powerful support for accessing the information space at both the entity level and the overview level.

We regard flexibility, extensibility, and ease of use as important issues to consider while constructing a rich information structure. Flexibility concerns adapting the structure to various situations. For example, a single object or concept of the defining structure can be used for multiple purposes, adapting the structure
Go to contents 01

to different situations. As a concrete example, an object might denote either a single document or a hierarchy of document components. Extensibility relates to the fact that one is never able to envision all possible uses at the beginning of the development. As an example, the form of the precise document formats to be represented within the structure should not determine the information structure. From a representation of text and images, the application should be extendable to include simple line drawings without having to change the structure. Finally, ease of use concerns among others the semantics of the data, e.g., that the semantics can be extended or adapted easily without modifying the overall structure or application. Users should be able to build or adapt the semantics of the data during the process of building the information structure. This requires the semantics of the data not to be encoded in the structure.

We propose two techniques for dealing with these issues in the construction of a rich information structure: augmenting the structure’s relatedness, and the separation of syntax and semantics. Both techniques relate to all three issues, and are themselves closely connected. Augmenting the structure’s relatedness increases the number of information entities and their relationships. One approach to this is the decomposition of documents into components within the context of their content matter. This ensures that the created document components are automatically related in a compositional manner and through content relationships. The semantics of how this decomposition is done does not affect the structure from a representational point of view.

The separation of syntax and semantics supports the independence of the structure from encoded semantics information. The semantics of the document decomposition can be described in a virtually separate structure. This “semantics structure” can be generic and flexible from a representational point of view. It can be seen, for example, as a network of keywords that describe concepts or aspects of the documents and their components. Relationships that are specified between these and the document entities form the basis of the information structure. Relationships between semantic entities automatically enrich and tighten the network of relationships within this information structure. Since the structure of the semantic entities is flexible, it is easy for users to extend and adapt both the entities and the structure at any time.

In this paper we discuss and elaborate the ideas and methodology presented here, and we describe an exemplary application that we have developed for the purpose of presenting architectural analyses.

Complexity of information structures

Considering abstractions as document entities in an information environment on the web, image archives, architectural analysis presentation tools, and electronic document management systems are all examples of environments for storing and relating large collections of abstractions or documents. These documents can be of various formats. While the main uses of these various environments are distinct, as archiving, presentation, and collaboration tool, respectively, these are all concerned with the complexity of organizing an information space composed of a large number of information entities and their relationships. The main question is how to organize the information and to relate the individual entities within this organization in order to support searching and browsing the resulting information structure. Specifically, there is a need for an information organization that enables an outsider to access this information effectively, independent of the viewpoint of the person who conceived it.

A document-based approach, treating the individual documents as entities or objects that are organized and related according to different categories and attributes, offers a flexible organizational framework (Tunçer and Stouffs, 2000). This approach has both advantages and disadvantages. Since the semantics of the data is not encoded in the structure,
it avoids a rigid organizational structure. This makes it flexible to input and organize the documents. However, the resulting information structure defined by the documents and their relationships is rather sparse. This loose structure offers little support for searching and browsing.

Our goal is to achieve a rich information structure that offers the flexibility of a document-based approach, though within a powerful organizational framework. A rich information structure, consisting of a tight network of information entities and their relationships, provides increased value for accessing, viewing, and browsing this information. The most important aspect of this increased value is access to the information structure from alternative views than expressed by the individual abstractions. Especially in precedent-based learning, one is not only interested in the individual abstractions, but in an interpretation of the entire structure seeking information related to a concept of interest.

The best way to handle the complexity of architectural information is not through a simplification of the information structure. On the contrary, we advocate a complex information structure that enables views unbounded by the original abstractions. Using a common syntax to represent abstractions, these can be interpreted and related, and these relationships added to the representation. The result is an integrated structure of components and relationships, represented in a uniform way. The complexity of this information structure, however, should not stand in the way of its ease of use, especially when integrating individual abstractions into it. Therefore, the tools, mechanisms, and techniques for creating the integrated information structure should be as clear, straightforward, and intuitive to use as possible.

Two techniques for the construction of a rich information structure
In order to achieve a rich information structure, both the structure’s volume and its intra-relatedness must be augmented. For this purpose, we propose a decomposition of documents by content, and the separation of syntax and semantics.

Decomposition of documents by content
An electronic document management system offers a flexible framework for the organization of documents according to various categories and attributes. However, simply specifying one or more keywords for each document does not provide any information on the importance of a concept as specified by a document keyword, nor on the portion of the document this keyword applies to. Furthermore, users may opt to simply ignore keywords which apply to only part of a document. Instead, by allowing the user to select portions of a document for assigning keywords, many more keywords that better fit parts of documents can be specified and associated with the appropriate document portion.

An information structure can be extended by enabling the representation of selected document portions within the structure. This increases both the number of information entities and the network of relationships. Document components relate hierarchically to the original document entity. Assigning keywords to document components relates components between different documents. Such a decomposition of documents with respect to content information can be automated using keyword or concept recognition mechanisms for texts, image recognition mechanisms for images, shape recognition mechanisms for simple line drawings, etc. (Tunçer et al., 2001).

Separation of syntax and semantics
Another means to ensure flexibility in achieving a rich information structure is the separation of syntax and semantics. When the semantics of a document decomposition can be separately encoded, this semantic structure can augment the document structure without imposing a specific compositional structure. The form of this semantic structure may be dependent on the meaning it entails. It may be linear, such as in a chronological list of project phases, or hierarchical, as in a conceptual structure offering
various levels of detailing. It may also constitute a combination of various semantic structures, e.g., describing different aspects or parts of a typology. The structure’s complexity can be extended or reduced according to individual cases.

This separation of syntax and semantics offers extensibility and flexibility within a system without imposing a fixed frame of reference, as the semantics can easily be altered without an adaptation of the syntactic structure. Keywords can be imported as a network of concepts, organized according to their relationships and dependencies, then associated with documents. Since these keywords are themselves organized in a structure, the relationships between the components are augmented through the relationships between the keywords.

Rich information structures
The information structure created through this approach offers new possibilities for accessing, viewing, and interpreting this information. First, it allows one to access specific information directly instead of requiring a traversal of the document hierarchy. Individual components can be reached and retrieved more quickly when provided with more relationships. Second, components can be considered from a different point of view. The location of a component in the structure is no longer only defined by its place in the document hierarchy. Instead, components provide direct access to other related components, forming a part of the first component’s view. Third, one can access the information structure from alternative views to those that are expressed by the individual design documents. New compositions of components and relationships offer new interpretations of the structure and generate views not inherent in the structure as created by the original documents. (Tunçer and Stouffs, 2000). The specifics of the semantics and the depth or level of the decomposition are left up to the user, when integrating an abstraction into the overall structure.

Flexibility, extensibility, and ease of use
We discuss the effects of both techniques on the issues of flexibility, extensibility, and ease of use mentioned in the introduction, with respect to the construction of a rich information structure.

Ease of use
We consider a structural decomposition of a document as opposed to a semantic one, that is, components are defined as subsets of the overall document and using the same representation. This approach to decomposing documents provides a uniform structure that is easily adaptable, unlike a semantic decomposition according to a product model. In this structure, the semantics of the decomposition is separately specified by a categorization of the components. Separating semantics from syntax in the decomposition offers important flexibility in constructing a rich information structure. Users can alter either the decomposition or the categorization without affecting the other. Furthermore, the user has full control on the effective positioning of his/her document within the categorical organization, i.e., by selecting the number of keywords assigned and the level of decomposition. Additionally, the semantic structure can be incorporated from an external framework or specifically defined corresponding to the project. The latter may require the structure to be constructed across the viewpoints of different groups or users. As a result of the separation of syntax and semantics, this construction can easily be achieved, and altered even after documents have been decomposed.

Extensibility
We consider the concept of a rich information structure independent of the types and formats of the documents or abstractions that define this structure. Though, at this time, we are particularly interested in abstractions in the form of texts, images, and simple line drawings in order to explore and apply this concept, at the same time, we do not want to exclude
the extensibility of any implementation to abstractions in other forms.

XML (eXtensible Markup Language) (http://www.w3.org/TR/REC-xml: Oct 2000) is particularly suited for the purpose of describing a decomposition of documents related with keywords. XML is a meta-language that serves to define markup languages for specific purposes. By specifying a grammatical structure of markup tags and their composition, a markup language is defined that can be shared with others. One of the strengths of XML for this purpose is its ability to represent information structures: how various pieces of information relate to one another. Once a structure is agreed upon, decompositions of existing documents can easily be expressed in XML.

Using tools for scanning texts and images and recognizing keywords, concepts or patterns, such conversion can be automated. The XML structure ensures that the data is consistently organized and is both machine- and human-readable. Furthermore, if the structures agree, XML documents can be plugged into a larger context.

**Flexibility**

XML also serves to integrate such a decomposition of documents into an existing web-based EDM environment. When decomposing documents in XML, the effect of this decomposition on the structure and representation of the EDMS can be kept to a minimum. Rather than having to replace a document entity by its composition hierarchy of document components, the XML decomposition can be linked to the document as an attribute, simply as text. By interpreting this document attribute, the decomposed document structure can be retrieved and presented. In this way, both the flexibility and the effectiveness of the EDMS is improved without altering the structure of the EDMS, nor imposing any fixed frame of reference.

**Implementation**

We are developing an application that will combine the described techniques in the form of a web-based tool for the presentation of architectural analyses in an educational setting. Analysis plays an important role in design, research, and education. In education, it is common practice for architecture students to collect abstractions on prominent buildings relevant to their design task in the early stage of design. As a result, the study of important historical precedents or designs plays an important role in design instruction and in the students’ design processes (Akin et al., 1997). Students also benefit from a collaboration with peers, in which they form groups to do an analysis of various aspects of a building or a group of buildings. By integrating the respective results into a common, extensible, library, students can draw upon other results for comparisons and relationships between different aspects or buildings. The web offers many examples of environments to build up, store, and present architectural analyses on a wide variety of subjects (e.g., Madrazo, 2000). Commonly, these analyses consist of a collection of abstractions describing different aspects of the building. These abstractions exist in a variety of formats. An information structure that integrates the different aspects of the analysis, such that the analysis can be interpreted and used in ways other than the original abstractions present, would be particularly useful in education. This organization can be augmented by applying the methodology presented in this paper.

The analysis presentation tool allows for a decomposition of documents by content using a hierarchical typology structure. The input to the application is a set of design documents in the form of images, texts, and simple line drawings, and a type hierarchy. The output is an integrated structure of components and relationships. In between, a number of steps are traversed: documents are broken up into their components, and these components within and between documents are related through types. We are using XML for the purpose of decomposing documents and integrating these into a single structure.

The interface allows the user to view both the type and document hierarchies and their relationships in an intuitive way. The component view presents a
component or type together with its immediate neighbors within the hierarchy, and displays all other components that share a type with it. This allows one to browse the structure and interpret relationships, and as such lets the user be guided to interesting overviews. Types mainly serve as binding elements in the structure providing semantic relationships between components. While a component’s types are presented as properties of the component, its relationships are given as component-to-component relationships. This not only ensures that the links are presented as shortly as possible, tightening the information structure, but it also shifts the focus onto the content, rather than on the structure surrounding it. Types further serve a role as index to the information structure. Access to the analysis is provided through the collection of abstractions and from the type hierarchy.

**Conclusion**

Complexity is a necessary characteristic of information models if they are intended to yield more than a few predefined viewpoints to the information. Targeting a largely unfamiliar audience, the indeterminacy of viewpoints provides the possibility to anticipate individual requests from the audience. Unexpected viewpoints derived from the information can also invoke new interpretations of existing information, which in turn can lead to creative discoveries. We argue that by decomposing documents by content and integrating these into a tight structure, such complexity can be achieved in a simple approach. This methodology implies both expanding the document structure, replacing document entities by detailed substructures, and augmenting the structure’s relatedness with content information. The result is a rich information structure that can easily be constructed from a collection of common abstractions and, through its complexity, presents new ways of accessing, viewing, and interpreting this information. Furthermore, this methodology offers flexibility, extensibility, and ease of use in the construction of the rich information structure.

**References**


