Hans-Georg Fill

Transitions Between Syntax and Semantics Through Visualization

In the context of processing information the transition from syntax to semantics and vice versa has been studied in several fields and for a variety of purposes. One area where it is of particular interest is the domain of law and legal informatics where these transitions play an important role for supporting human communication and understanding as well as for enabling machine processing. In this paper we discuss how the techniques of visualization can be used for supporting such transitions by abstracting information to a meta level. We illustrate this by using selected examples and derive directions for further research.

Category: Articles
Region: Austria

Contents

1   Introduction

[Rz 1] The increasing amounts and the complexity of information that we are confronted with today as well as the speed in which information changes, requires the development of innovative approaches for handling these challenges. Examples where it is particularly necessary to engage in corresponding research activities can be found in such diverse fields as the medical and healthcare domain where the global research activities on the development of new drugs and therapies lead to massive amounts of valuable information for curing diseases, or in the domain of business where ungraspable amounts of complex financial transactions are executed and need to be supervised and checked for their compliance to regulations. Also in the domain of law and legal informatics, the complexity and volume of legal texts and professional publications require new modes of interaction and in many cases the use of specialized technological components cf. [KIENREICH et al., 2012], [STÖGER-FRANK, 2012].

[Rz 2] In order to support humans in identifying relevant information several fields have made important contributions. Not only has the field of computer science fundamentally enabled us to let machines take over highly repetitive tasks to filter, select, and process high volumes of information, and share them worldwide within milliseconds. Also the field of visualization has developed methods and techniques to support humans in understanding the content of complex pieces of information and for gaining value. These concrete applications are however preceded by abstract meta levels and abstract terms and ideas, which Lachmayer and Schweighofer denote as «universalia ante rem» [SCHWEIGHOFER AND LACHMAYER, 1997]. In this context Lachmayer and Schweighofer view «abstraction» as a procedure of complexity reduction to identify essential aspects.

[Rz 3] When we therefore investigate the possibilities for supporting humans in interacting with complex information and aim to develop insights into the inner workings of the underlying mechanisms, one has to consider two fundamental aspects: a. how information is structured and b. how the meaning of information is conveyed. As we will discuss in the following, the structure of information can also be related to syntax and the meaning of information to semantics. To support humans in processing information — not only in terms of handling information but also in terms of understanding its content and deriving according actions — we can identify two possible directions. Firstly, we can provide support from a technical perspective, i.e. using modern computer systems that take the burden from us to process information ourselves in large amounts. Or, secondly, we can provide means of communication, i.e. in the form of aids that are developed to better understand complex relationships by explicating important and relevant aspects. Although these two directions can be pursued independently they can also be joined, e.g. by developing IT-based means for communication or communication means for enabling machine based processing.

[Rz 4] In all these cases, the syntax and semantics of the underlying information have to be considered. When developing new approaches one therefore has to decide whether to start with a

focus on the syntactic, i.e. structural aspects of information, or with the semantic aspects, i.e. the meaning and content of the information. When it comes to solutions that involve machine based processing, both syntactic and semantic aspects need to be mapped to formal syntactic constructs that in the end correspond to the basic functions of the target machine. For human processing this mapping takes place implicitly by assuming a certain amount of knowledge about a domain. These tasks are the basis for a variety of applications that have been developed e.g. in the context of: natural language processing, where textual syntax and semantics are transformed into machine processable statements; knowledge engineering, where knowledge is made explicit in the form of machine processable information such as formal ontologies; or conceptual modeling where knowledge is represented in information formats for human communication and understanding.

[Rz 5] In the following we will at first briefly discuss some foundations of syntax, semantics and visualization. We will then describe two ways for transitioning between syntax and semantics through the use of visualizations. Finally we will discuss the implications of these approaches and directions for further research.

2 Foundations

[Rz 6] In the following we briefly outline characterizations of syntax and semantics to achieve a common understanding of these concepts. According to Zemanek syntax is concerned with the ordering of signs and the rules that lead to this ordering, whereby the meaning of the signs is not taken into account cf. [Zemanek, 1992 p.73]. Syntax can therefore also be related to the *form* of information that is able to convey content and fundamentally enables humans and machines to process information by defining the structure in which information has to be presented. Although this characterization states that syntax is not concerned with the meaning of the signs, it can be a hard task in practice to achieve this independence to a full extent. Unless a formal language such as mathematical notation is used, many signs already carry some meaning, e.g. due to cultural conditioning or personal experience.

[Rz 7] Regarding semantics, which is concerned with assigning meaning to signs and their combinations, it needs to be detailed where this meaning originates from. We will therefore take up a view that is both common in semiotics as well as in computer science. This view as e.g. described in [Nöth, 2000 p.91] in reference to Charles W. Morris and also in [Zemanek, 1992 p.74], focuses on semantics as a *reference relation* between entities for the purpose of describing one entity through the reference to another. The combination of this view with the above characterization of syntax then leads to a description of the content of information, i.e. the meaning that is carried by a sign or statements based on signs. Thus, the content that is encapsulated in the form is interpreted via a reference relation to some other entity. On a formal level this reference can also be defined via mappings between formal constructs where a «definiens» is mapped to a «definiendum» to describe its meaning, cf. [Messer, 1999]. Special cases in this context are *formal languages* as artificially constructed languages where the meaning of every sign is unambiguously defined by its form cf. [Tarski, 1936 pp.268]. In these cases it is assured that no a-priori references exist for the signs of such a language, which can in practice only be achieved by highly abstract mathematical notations. Another important aspect in regard to semantics is that of *context*. The signs, the reference relations and their targets are part of a certain context which may also be denoted as a «domain» of discourse or similarly a «model» in the area of model theoretic semantics. If this context has boundaries we can refer to a concept that researchers in the area of databases and

logics denote as the «closed-world-assumption» — in contrast to the «open-world-assumption» where the context is left open.

[Rz 8] Also in the field of visualization the concepts of syntax and semantics are used. For a characterization of visualization we will revert to a definition elaborated by the author, where it is defined as «the use of graphical representations to amplify human cognition» [FILL, 2009 p. 19]. Similarly to the above characterizations, visualizations are composed of fundamental entities that can be structured and assembled to create more complex entities. One well-known approach for defining the basic entities for visualizations has been made early by the primitives of geometry as described by Euclid cf. [JOYCE, 1997]. Although these descriptions only focus on shapes — i.e. based on points as the most fundamental primitive, then lines, circles, rectilinear figures, trilateral figures and so on — they already come very close to what we have stated above for syntax. The spectrum of the visual primitive concepts can be extended to visual variables such as size, color, brightness, orientation, texture and the planar variables horizontal and vertical position cf. [BERTIN, 1983 cited after MOODY, 2009]. In order to express semantics, references from visual primitives to formal constructs or also informal textual statements can be established. For example, Ware lists thirteen fundamental graphical codes for node-link diagrams and their commonly associated meanings by using textual descriptions [WARE, 2000 p. 226]. In [FILL, 2009] formal syntactic descriptions of visualizations in the form of visual objects are linked via reference relations to formal statements in ontologies. Thereby not only the static meanings associated to the visual objects can be assigned, but also the dynamic behavior of these objects in terms of their embedded control structures and the processing of variables is made explicit on a formal level.

3 Syntax — Semantic Transitions through Visualization

[Rz 9] With these foundations we can now investigate two paths for transitioning between syntax and semantics and show how visualization can be used to support these steps. The goal thereby is to make information from a knowledge domain processable either by humans or by machines. The first path starts with the consideration of a «knowledge domain». This knowledge domain defines the context of the further steps and acts as the basis for the information that is to be processed. The next step concerns the «identification of syntax structures». This involves the identification of the signs as the primitive entities and the rules for the combination to express valid statements. This step may also involve a certain degree of abstraction: sometimes the original information stemming from the knowledge domain shall or can not be represented directly in the syntax structures but needs to be taken to a more abstract meta level in order to meet the goal of being useful in processing. An example where no such abstraction is present would be natural language processing where it is aimed for a highly exact processing of the original information pieces. In contrast, when complex information needs to be made understandable for humans, e.g. to explain complex legal processes, abstraction from the underlying legal texts may be necessary to highlight the relevant parts of the process structure.

Figure 1: Transitioning from syntax to semantics

[Rz 10] Subsequently, when the syntax structures were identified, they can be «enriched with the semantics» of the information content. This involves the provision of references to either other formal or informal constructs to assign meaning to the signs and statements of the syntax. Again, as in the first step also abstractions for the original semantic references may be necessary to achieve proper understanding. As mentioned above, it also needs to be taken into account if the result of the transitions shall be processed by a machine. Then, the enrichment with semantics has to be conducted such that every semantic reference and every target of the semantic reference can be finally mapped to machine executable functions.

[Rz 11] To illustrate how visualization can participate in this transitioning process and act as an application of these meta level concepts, we will revert in the following to an example from Friedrich Lachmayer from the domain of legislative workflows. As

Figure 2: Legislative Workflow, Source: [Lachmayer, 2008]
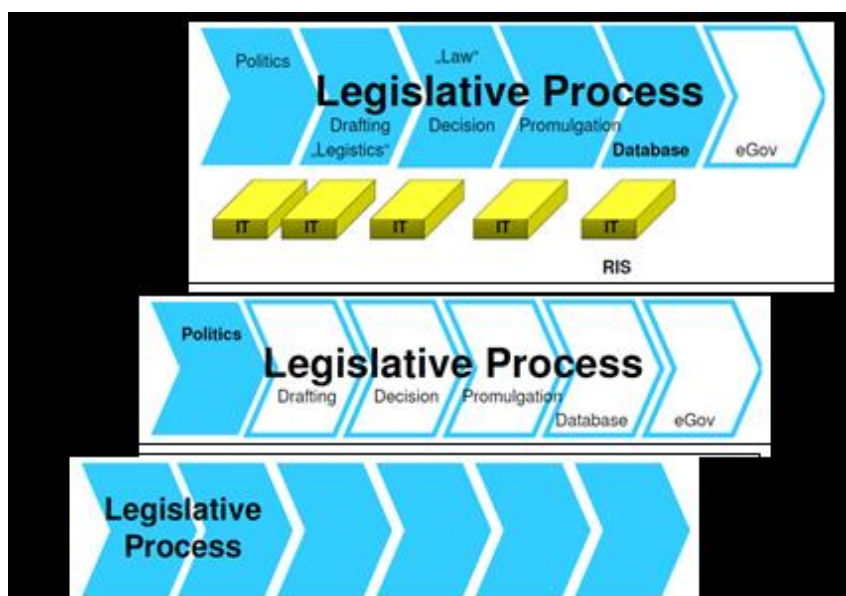
[Rz 12] shown by the excerpts of the original document in figure 2, the visualization consists of three stages. At first, a basic process like structure is presented that already contains a textual label giving only a basic semantic reference. Although, based on a-priori knowledge the informed viewer is already provided with a vague idea about the meaning of the visualization, the primary focus rests on the syntax structure. The single symbols thereby indicate a sequence based on an intuitive understanding of the used signs. At the second stage (II), the structure is now populated with semantic references by adding labels to each of the process steps. This already permits a viewer to understand the content of the presented information. In (III) this is even further extended by adding additional labels to some parts, i.e. enhancing existing concepts with meta concepts such as «decision» —> «law». In addition, new symbols for showing the «IT» support of the process are added that already carry semantic references. One of these symbols is even further detailed by a concrete instance «Database» -> «IT» -> «RIS». In this way of using visualization, abstraction plays an important role as the fundamental structures of the process become directly visible and thus greatly adds to the understanding of information from this knowledge domain.

[Rz 13] The second path also starts from a «knowledge domain» — see figure 3. In contrast to the

first path however, the focus rests at the beginning on the «identification of semantic concepts». This may seem counter-intuitive at first as we have characterized semantics as reference relations from signs of the syntax to other entities. We can however also start with an investigation of these other entities that constitute the context of the information we want to process and then later add an appropriate processing structure. Therefore we first have to analyze the knowledge domain for the concepts that we want to represent, relate them to each other and optionally provide abstractions for the concepts. Then we design a syntactic structure for these concepts that permits to process them. Whereas in the first steps, the concepts may be arbitrarily positioned in relation to each other, the «syntax structuring» step involves the consideration of potential sequences that are necessary for processing.

Figure 3: Transitioning from semantics to syntax

[Rz 14] Again, we can illustrate how visualization can be used in this process. As shown in figure 4, we start in (I) with a so-called «visualizing analysis» based on a drawing by Colette Brunschwig that is well known in the area of legal visualization. It shows the legal aspects in a contract relationship. Although this drawing is complete in the sense that it contains both syntactic and semantic aspects to understand the situation, it would be hard for a machine to process this information. Despite available technologies in image processing, the underlying relationships between, for example, the actors in the pictures could not be directly interpreted by a machine. We can therefore view this drawing — from a machine perspective — as a collection of semantic concepts without any processable structure. In steps (II) and (III) we see how appropriate structures are added: first, by identifying the basic signs and relations between the signs. Second, the types of these signs are made explicit. Although we may have not yet progressed far enough with this syntactic structuring to achieve a useful processing apart from highlighting the concepts, the example shows the goal we are heading at, i.e. to transition from semantics to machine and human processable syntax.

Figure 4: Visualizing analysis for transitioning from semantics to syntax, Sources: [Brunschwig, 2011], [Fill, 2013]
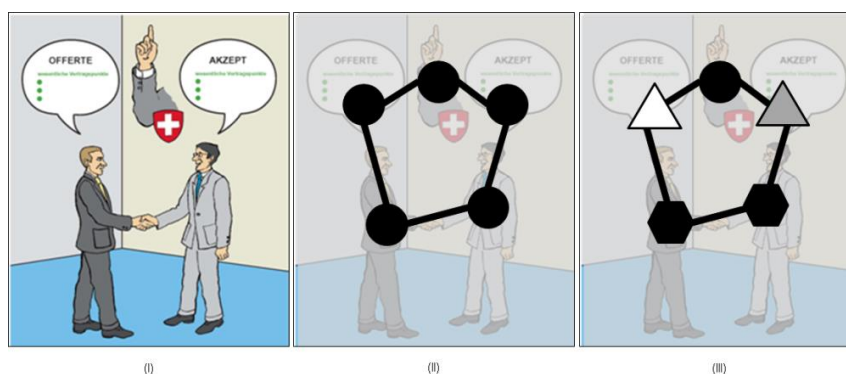
4   Implications and Further Research Directions

[Rz 15] Based on these two ways of transitioning between syntax and semantics using visualizations we gained insights on these fundamental aspects of enabling information processing for humans and machines. It thereby became apparent that not only the differences between syntactic and semantic aspects can be very effectively conveyed by using visualizations. It could also be shown that abstractions of concepts can be easily communicated by using visual representations. Doing something similar using for example formal mathematical notation is certainly possible and has been done many times but would require in-depth mathematical knowledge to be as easily processable. It thus seems worthwhile to engage in further research in these directions. In particular it needs to be further investigated how the cognitive and to a large degree also very creative process of transitioning from a knowledge domain to either syntactic structures or the context in the semantic sense can be better supported. Although a large number of examples have been elaborated in the past for these cases, it still needs to be made explicit, for which aspects and dimensions visualization can effectively support these transitions. Potential starting points for such analyzes could be: visualizing analyses using conceptual models and ontologies, formal visual definitions of syntax, semantics and processing algorithms or combinations of cognitive and information/data oriented visualizations for enabling transitions.

5   References

Bertin, *J.*, Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, 1983.

Brunschwig, *C. R.*, Multisensory Law and Legal Informatics — A Comparison of How these Legal Disciplines Relate to Visual Law. in: Anton Geist, Colette Brunschwig, Friedrich Lachmayer, F. & Günther Schefbeck (Eds.), Strukturierung der Juristischen Semantik — Structuring Legal Semantics, Festschrift für Erich Schweighofer, Editions Weblaw, Bern, pp. 573-668, 2011.

Fill, *Hans-Georg*, Presentation on: Polysyntactic Meta Modeling: Historical Roots in the Work of Raimundus Lullus, Accompanying text in: Erich Schweighofer, Franz Kummer, Walter Hötzendorfer (Eds.); Abstraktion und Applikation — Tagungsband des 16. Internationalen Rechtsinformatik Symposiums, books@ocg.at, Band 292, Wien, 2013, pp. 439-444.

Fill, *Hans-Georg*, Visualisation for Semantic Information Systems, Gabler, 2009.

Joye, *D.E.*: Euclid's Elements, Clark University, URL: http://aleph0.clarku.edu//java/elements/toc.html (access 15-03-2013), 1997.

Kienreich, *Wolfgang/Lex*, *Elisabeth/Rapp*, *Stefan*, Maschinelle Lernverfahren für die automatische Klassifikation von juristischen Dokumenten, in: Erich Schweighofer, Franz Kummer, Walter Hötzendorfer (Eds.); Transformation juristischer Sprachen — Tagungsband des 15. Internationalen Rechtsinformatik Symposiums, books@ocg.at, Band 288, Wien, 2012, pp. 83-87.

Lachmayer, *Friedrich*, Austrian Legal Information System RIS 02, Legislative Workflow, URL: http://www.legalvisualization.com/download.php?hash=e69efbcdd1f3ec9c3df40ce8eec981ef?fname=Austrian_Legal_Information_System_RIS_02%2C_Legislative_Workflow.pdf (accessed 10-03-2013), 2008.

Schweighofer, *Erich/Lachmayer*, *Friedrich*, Ideas, Visualisations and Ontologies, in: P. Visser, R.G.F. Winkels, Proceedings First International Workshop on Legal Ontologies, LEGONT'97, Juli 1997,

Melbourne, Victoria, Australia, 7-13.

Messer, *B.*, Zur Interpretation formaler Geschäftsprozess- und Workflow-Modelle, in: Jörg Becker, Wolfgang König, Reinhard Schütte, Oliver Wendt and Stephan Zelewski, Wirtschaftsinformatik und Wissenschaftstheorie — Bestandaufnahme und Perspektiven, Gabler, 1999, pp. 95-123.

Moody, *D. L.*, The «Physics» of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering, IEEE Transactions on Software Engineering 35(6), 2009, pp. 765 ff.

Nöth, *Winfried*, Handbuch der Semiotik, Metzler, 2000.

Stöger-Frank, *Angela*, Aus drei mach eins: Eine Suchmaske für drei Datenbanken (FINDOK, LE-XISNEXIS, LINDE) zur Reduktion des Suchaufwandes, in: Erich Schweighofer, Franz Kummer, Walter Hötzendorfer (Eds.); Transformation juristischer Sprachen — Tagungsband des 15. Internationalen Rechtsinformatik Symposiums, books@ocg.at, Band 288, Wien, 2012, pp. 53-55.

Tarski, *A.*, Der Wahrheitsbegriff in den formalisierten Sprachen, Studia Philosophica I, 1936, pp. 261-405.

Ware, *Colin*, Information Visualization — Perception for design, Morgan Kaufman, 2000.

Zemanek, *Heinz*, Das geistige Umfeld der Informationstechnik, Springer-Verlag, Berlin, Heidelberg, 1992.

Hans-Georg Fill, Assistant professor, Research Group Knowledge Engineering, University of Vienna.