Meta data driven legislative information retrieval based on semantic and pragmatic knowledge models

Valentina Bartalesi Lenzi¹, Carlo Biagioli², Amedeo Cappelli¹⁻³, Rachele Sprugnoli¹, Fabrizio Turchi²

¹ CELCT, Via Sommarive 18, 38100 Trento, Italy
{bartalesi, cappelli, sprugnoli}@celct.it
² ITTIG-CNR, Via dei Barucci 20, 50127 Firenze, Italy
{biagioli, fabrizio.turchi}@ittig.cnr.it
³ ISTI-CNR, Via Moruzzi 1, 56124 Pisa, Italy
{amedeo.cappelli}@isti.cnr.it

1 Functional Framework: towards a formal provision theory

The Italian NiR project has already promoted the standardization of legislative data and its annotation by means of meta data on acts and their contents, accepting our suggestion to add to the description of the textual elements (partitions), the description also of their regulative functions (Biagioli 97).

Semantics of a text is the result both of its atomic components (simple and complex terms), of the meaningful aggregations of such components (simple and complex clauses, which might be likened to linguistic acts, i.e. to normative micro-acts) and of the text as a whole (the normative macro-act).

The functional elements suggested are called dispositions or rules and are intended as typical legislative statements. The formal profile has given place to DTD-NiR, while the functional to the analytical internal meta data.

Although no linear relationship can be found between surface and functional patterns in a text, there is a consistent correspondence between text atomic items (paragraphs) and meaningfull/functional units (provisions), intended as normative speech acts. Based on this finding, it can be reasonably argued that this is no casual correspondence: indeed, each provision is, from a general standpoint, the meaning of each basic text unit.

Functional structure includes classes and types of provisions and arguments of provisions. This classification has been derived from legislative techniques and practices, by taking also into account, where necessary, of the best-known theories on norms.

Dispositions are divided into two main class, changes and rules split those into the two major categories that are the main subject of normative theories, often referred to as constitutive and regulatory rules, according to the well-known distinction drawn by J. Rawls (Rawls 1955): justifying a practice and justifying a

¹ Or rhetoric and illocutory, according to Branting (Branting, Lester, 96).
particular action falling under it, where practice stands for any form of activity specified by a system of rules which defines offices, roles, moves, penalties, defenses, and so on, and which gives the activity its structure.

Class also can be used in the semantic mark up of complex partitions, having a more general meaning than dispositions.

Arguments are the main focus of provisions and give them their logical structure. Its own arguments are always to be found in a provision of a certain type, to be regarded as such (see, for instance, the definiendum for a definition, or the addressee for an obligation). Arguments may be either single or multiple and appear in the texts in various forms: explicit direct, explicit indirect, explicit variable (unknown) and implicit. (ex.)

In addition to the type of provision, arguments contents typically make up the information of section/part titles in Italian legislative texts: law maker uses provision typology and its arguments to inform about the ratio of the following partition, at list in the Italian legislative drafting practice.

In the end the legislative text has been deeply modelled as a set of provisions, while each provision has been modelled as a framework made of several well known internal components said arguments (slots). The contents of every argument are keywords that belong to a domain controlled vocabulary, organised into classes, as explained below.

A double modeling approach has been in fact chosen. Provisions D-A model express the pragmatic side, how legislator organize and express rules about domain entities, while domain entities are classified and described ontologically.

The double modelling of the pragmatic and semantic profiles of texts meanings, is in tune with the Breuker - Hoekstra statement an ontology is in the first place a set of terminological definitions built around a taxonomic back- bone, while a framework is a an assembly of concepts or types of knowledge that reflect recurrent patterns of use (Breuker 2004).

1.1 Formalization of Disposition Theory

The formal provision theory developed is a terminology (in the technical sense of description logic) for classifying fragments of legal texts. The domain of discourse is the set of textual fragments, that is to say, the formalization talks about textual fragments and it considers them as instances of meta-data, i.e., the dispositions structured in the terminology. The terminology of dispositions presented here can be seen, therefore, as a theory of a possible set of meta-data that can be used for the mark-up of legal texts. Meta-data are isolated a priori and the logical relations holding between them are imposed axiomatically. As such, the terminology of disposition presented cannot be properly considered a theory of normative concepts. The formalization of Biagioli’s theory of provisions, amounts to a logic of dispositions and it can be viewed as an attempt towards a formal foundation of legistics (i.e. the set of techniques for legal drafting). Description logic has been used: S III F (D) which is a notational variant of OWL(Web Ontology Language)-lite. OWL is a standard knowledge representation language within the Semantic Web community which enjoys appealing computational properties (worst case EXP-time).
formalism we propose is of a modular nature. We will first provide a sort of minimal theory of dispositions essentially corresponding to a taxonomy. Stronger logical relations could later on be imposed in order to enable richer inference patterns and to better characterize the dispositions at issue. (Biagioli-Grossi 2007)

For their effective employment arguments are to be synthetized and connected to structured dictionaries and terminological collections. In practice the model allows to evidence the terms of the dispositions that carry out a particular role, expliciting therefore theirs meaning in that context (disposition). The information on the general meaning of the term will be found in terminological collections.

The LME\textsuperscript{2} system works with every classification of domain relevant keywords, from flat glossary to deeply structured ontology. More complex is the description of domain entities, more easy will be to express complex contents descriptions in the searching system queries (metaSearch). To test the LME system, a domain ontology has been developed, concerning the personal data protection Italian law.

### 2 The Role of Ontology: ontoPrivacy

Laws contain provisions which deal with domain common sense entities (arguments contents), but they do not provide any general information on them, except the case of definitional provisions. The use of an ontology allows the possibility to acquire such additional general information. Moreover an ontology allows to obtain a normalized form of the terms by which entities are expressed, so that they can be indexed and used in the analytical meta data querying process of law document search and retrieval.

In recent years there have been several initiatives for the development of legal ontologies in order to offer a solid support for the acquisition, sharing and re-use of legal knowledge. In particular, we have taken into consideration two core ontologies: 1) LKIF Core Ontology (Breuker et al., 2007): it is inspired by the LRI Core ontology (Breuker, 2004) effort and it was developed within the Estrella project\textsuperscript{3}. The Legal Knowledge Interchange Format ontology consists of 14 modules covering both general concepts, like place and time, and more specific ones, like intensions and norms. 2) Core Legal Ontology (Gangemi et al., 2005): it organizes juridical concepts in classes of entities that are specializations of the DOLCE+ foundational ontology (Gangemi and Mika, 2003). The three basic categories are Region (space and time), Endurant (object or substance, either physical or not), and Perdurant (events, states, or processes).

At the same time, we have taken the Legal-RDF Vocabularies as a reference. These Vocabularies have been developed by John McClure, Legal-RDF\textsuperscript{4} director, for the annotation of narrative legal documents in the Semantic Web perspective. These documents are organized according to the classes and properties defined by the W3C s Resource Description Framework (RDF) and Web Ontology Language

\textsuperscript{2} Law Making Environment is the whole project of legislative drafting support, through a set of software tools and related knowledge management techniques.

\textsuperscript{3} http://www.estrellaproject.org/index.php/Main_Page

\textsuperscript{4} http://www.legal-rdf.org
The goal of this modeling is to reflect the real world to the maximum extent possible. Seven classes are defined: Actors (individuals and groups), Dramas (events, divided into discrete actions and open-ended activities), Props (products and legal properties), Roles (participant, occupations, and legal and family roles), Scenes (place and time), Scripts (document types), and Themes (topics of a script or drama).

Our work focused on the study of the ontologies mentioned above in order to organize into classes the instances of a glossary of keywords extracted from the Italian Personal Data Protection Code\(^5\) (d.lgs. 196/2003). Therefore, ontoPrivacy can be considered a lightweight ontology consisting in a set of concepts and hierarchical relationships among them and not including axioms and constraints, peculiar of the heavyweight ontologies (Corcho et al., 2003). Moreover, following the Visser definition (Visser and Bench Capon, 1999), ontoPrivacy can be defined as a statute-specific ontology because it has been developed for a specific sub-domain of the Italian law.

The glossary, base for our experiment, has been manually created by an expert in the legal field and it is made up of both specific terms of the Public Administration domain (e.g. *atto amministrativo*/administrative act) and generic words (e.g. *razza*/race). Synonymy, hypernymy and hyponymy relations are identified among such terms (e.g. administrative act and judicial act are kinds of act).

Taking the relevant terms contained in the glossary as starting points, we have followed a bottom-up approach to create the ontology. Vocabulary and ontology are, in fact, closely tied with a two-way relation: a lexicon can be the basis for a well-built ontology and an ontology can serve as foundation to lexicon organization (Hirst, 2003).

The classes in ontoPrivacy have been defined by analyzing the lexical entries of our glossary and the relations among them. In Figure 2.1 we show a tree where the upper classes of ontoPrivacy are presented: each box represents a different concept, where *Thing* is the universal class that includes everything which is identified. The *is-a* arrows relate a subclass with its superclass.

![Fig. 2.1: The upper part of ontoPrivacy](http://www.parlamento.it/parlam/leggi/deleghe/03196dl.htm)

ontoPrivacy has six main classes: (i) Event: something that happens at a given place and time; (ii) Scene: the place and time where an event occurs; (iii) Artificial Object: are objects (i.e. artifacts) intentionally made to serve a given purpose; (iv) Mental Object: non-physical objects, the things we mentally manipulate like conceptualizations; (v) Legal Entity: as defined in 1999/65/CE, any natural person or any legal person, provided that it has been established under Community law or the

\(^5\) [http://www.parlamento.it/parlam/leggi/deleghe/03196dl.htm](http://www.parlamento.it/parlam/leggi/deleghe/03196dl.htm)
applicable national law and has been given legal personality or has the capacity, in its own name, to hold rights and obligations of all kinds, to conclude contracts and to be a party to legal proceedings; (vi) Role: functions played by individuals and groups.

A detailed description of each ontoPrivacy subclass is reported below together with examples and explanatory figures.

Event
An event is a situation that happens or occurs, involving participants and it can be described as a change of state.

There are many Event subclasses. Each Event subclass is named by a noun-phrase, e.g., AdoptionEvent, AgreementEvent, IdentificationEvent and so on. One or more keywords are assigned to each subclass: for example, *decisione*/decision and *deliberazione*/resolution are two instances of the DecisionEvent subclass.

Scene
The Scene class defines places and times.

A Place may be either a Boundary, Complex, Space, Structure, or Surface (e.g. *paese extraeuropeo*/non European state, *Repubblica italiana*/Italian Republic). A Temporal Expression (TE) can be a duration (e.g. *tre anni*/three years), a point (e.g. *17 Luglio 1999*/July 17th 1999), an interval or a set of times (e.g. *ogni mese*/every month).

At present, there are no temporal expressions in our glossary but we intend to use an automatic annotation tool to recognize and normalize a broad variety of TEs which can be found within an input text.

Artificial Object
Artificial Objects can be tools, machines, pieces of furniture, and many other things made by human beings. They are contrasted with natural objects like trees and animals. For example, in our ontology *fibra ottica*/optical mean, *cavo*/wire and *dispositivo elettronico*/electronic device are instances of the Artificial Object class.

Mental Object
The Mental Object class is divided in two subclasses: Information Object (IO) and Conceptualization.

An Information Object is a reification of pure information as described in Shannon s communication theory. An IO has the following properties: a support that realizes IO, one or more combinatorial structure (or code), according to which IO is ordered, a meaning that IO expresses, a reference that IO is about, and one or more agents that interpret IO (Gangemi et al., 2005). For example the Italian Personal Data Protection Code is an IO, expressed in contemporary standard Italian, realized by e.g. a paper copy, with a related meaning as interpreted by an agent with an average knowledge of the Italian law, and it is about judge-ruled process.

Information Object is further categorized in Juridical Information Object and General Information Object. A Juridical IO has a context and a content related to the law. It is defined at least partly by reference to its effect upon juridical relationship. There are three type of Juridical IO:

1. Administrative IO, e.g. *atto amministrativo*/administrative act;
2. Judiciary IO e.g. *sentenza*/judgment;
3. Legislative IO e.g. *provvedimento* / provision.

A General IO is not related to the legal field. E.g. *lettera raccomandata* / registered letter, *posta elettronica* / e-mail.

Figure 2.2 shows the Information Object class together with its dependent elements.

For what concerns the Conceptualization class, we consider concepts as formed and controlled by mental processes that transform or transfer them. Instances of the Conceptualization class are, for example, *licitità* / lawfulness and *esattezza* / accuracy.

**Legal Entity**

The Legal Entity class is divided into 2 subclasses: Natural Person and Legal Person.

A Natural Person, as defined in Wikipedia, is a human being perceptible through the senses and subject to physical laws, as opposed to a Legal Person. A Legal Person is a group of organizations or persons which acts as a single individual, according to the law.

A Legal Entity always plays a role, that is a function performed by a certain entity, when it enters in relationships with other entities.

**Role**

A Role is a function played by a Legal Entity.

The Role class is divided in two subclasses: Social Role and Judiciary Role. Social Role is played by some agent in the context of social activities. E.g. *fornitore* / provider, *richiedente* / requesting party. This class is divided in: (i) Person Role: a role played by some natural person. E.g. *cittadino* / citizen, *lavoratore* / worker; (ii) Organization Role is a role which has a meaning in the context of an organization. I.e. the role defines the position of a natural person within the structure of an organization. E.g *ministro* / ministry; (iii) Legal Role: a role played by some legal person. E.g. *camera dei deputati* / data controller, *comune* / local municipality.

Judiciary Role is played by some agent in the context of legal activities. This class has a Professional Judiciary Role as subclass. Professional Judiciary Role is a legal profession of some person, examples: E.g. *pubblico ministero* / public prosecutor.

Figure 2.3 illustrates the Role class together with its dependent elements.
3 Semantic Markup: metaEdit

The main objective of our project is the Disposition-Argument model (D-A model) testing on a real legal domain, limited but meaningful; to such scope we have developed metaEdit, a web-based software program that allows the semantic mark up of drafted legal texts according to the following models: the provisions model and the concept model. The user can describe the semantic contents of each text partition first, qualifying its disposition type through the provisions model, thus capturing the basic intention of the legislator. Secondly, one can qualify the provision contents through the available concept model, thus describing to some extent the content and details of each provision. metaEdit allow the user, before getting involved in the semantic mark-up process, to edit, modify and also create a dictionary of concepts to be used in the semantic mark up, if it is not already available from the beginning. The metaEdit design is shortly shown in figure 3.1:
metaEdit allows the user the semantic mark up process, throughout the following steps:

from the legal act - in our case the Legislative Decree no. 196 of 30 June 2003 - we select the partition (article, or paragraph) to which you want to add the semantic data; naturally the mark up process is based on the associated XML legal act, according to the DTD devised within NiR project (DTD-NiR);

after having chosen the partition we are able to see its text, through a DOM parser that extracts the part of text from the XML document. At this point it's possible to assign a disposition type, such as obligation, prohibition, to the selected partition in the final step we assign values to arguments of the chosen disposition type catching the values from the legal vocabulary.

The semantic mark up includes the id partition, a unambiguous way to identify the partition of formal profile, that represents the gateway between the syntactic profile and the semantic profile. As example, we can see the following partition:

**article 22, paragraph 8:** Data disclosing health may not be disseminate

The simplified formal mark up, coming from the DTD-NiR, is the following:

```xml
<paragraph id="art22-par8" xml:lang="en">
  <num>8</num> <body>Data disclosing health may not be disseminate</body>
</paragraph>
```

In the semantic mark up process, the previous paragraph is seen as:

<table>
<thead>
<tr>
<th>partition</th>
<th>art22-par8</th>
</tr>
</thead>
<tbody>
<tr>
<td>disposition</td>
<td>prohibition</td>
</tr>
<tr>
<td>argument action</td>
<td>disseminate</td>
</tr>
<tr>
<td>argument object</td>
<td>data health .</td>
</tr>
</tbody>
</table>
Table 3.1: semantic mark up sample

When you assign the values to the arguments you rely on the ontological classes so that the available values belong to the appropriate ontological class, for example

<table>
<thead>
<tr>
<th>counterpart/addressee</th>
<th>Natural persons</th>
</tr>
</thead>
<tbody>
<tr>
<td>action/activity</td>
<td>Events</td>
</tr>
</tbody>
</table>

3.1 Mark Up Strategies: metaEdit

Mark up strategies can range between two extreme ends. Documentalist can decide only explicit entities to be marked, so adopting a low degree of text interpretation. Otherwise also implicit informations, detected tanks to an interpretative effort, can be added to the marked up entities. Obviously more complete is the mark up, more wide will be the answer in the searching for norms process.

The provisions belonging to the Change top class (Derogations and Modifications), are dispositions acting on the other provisions belonging to the Rule top class. Somehow they belong also to the nature (logical structure) of the acted provisions: a derogation to a permission provision is part of the whole permission, deriving from the connection of the derogated provision and the derogating one. Then the derogating provision has typical contents (arguments) of the derogated provision. Therefore derogations can be doubly marked up, as derogations and as provisions of the same type of the derogated provision.

4 Semantic Search: metaSearch

Legal documents (texts) show peculiar features and are quite different from other documents in terms of their fruition. Indeed, it can be argued that the reason underlying a considerable portion of documentary research on law collections is related to the need for identifying rules rather than law texts as such which are therefore, at the very most, an intermediate target.

The semantic mark up has been the starting point of metaSearch, an open source software tool to query the functional profile of legislative texts through D-A model and relied on the ontoPrivacy, the semantic domain ontology. This module allows the user to query a legislative information system on the base of two modalities of reasoning: a reasoning on provisions and their relations and/or on domain concepts and their relations. The metaSearch design is shown in Figure 4.1 and is built as a client server software:
Fig. 4.1: metaSearch design

Taking advantage of the knowledge deriving from ontoPrivacy, the searching software allows the user to formulate a better detailed query selecting a keyword from the structured vocabulary. Such legal vocabulary will not only be used in the query composition phase, but also in the presentation of the results: the final user will be able to widen his/her search by including terms which are in linguistic relation with the chosen one, through hyperonyms, hyponyms and synonyms relations.

The data exchange between the server and the client sides take place in a XML format through protocol HTTP. The Figure 4.2 shows the metaSearch main page:
Four simple steps are needed to create a query in *metaSearch*:

1. selecting the legal act having to be searched; it is possible to choose also the whole acts collection inherent the legal domain;
2. selecting the provision type: it's also possible to choose a class dispositions.; this selecting affects the arguments list showed in according to the D-A model. If the user choose a dispositions class the system shows the common argument among the dispositions belonging to that class;
3. selecting the specific argument; it is possible to choose the generic argument, any, to extend the query to all possible arguments of the disposition type chosen;
setting on the chosen argument by the legal vocabulary in two distinct manners:

- the first approach is selecting from a semi-flat terms list comprehensive of the hierarchical linguistic relations (hyponyms/hyperonyms/synonyms), showed in figure 4.3
- the second one is browsing and selecting from the ontological classes and roles created in ontoPrivacy, as appears in figure 4.4

Fig. 4.3: setting on the argument by legal vocabulary
The argument setting value process can involve more than one single term, generally tied to the previous ones through the OR logical operator or, if explicitly chosen, also the AND logical operator. Furthermore it's possible to choose a combination of OR and AND operators, and to manually add the needed parentheses in order to define a not ambiguous meaning to the final expression, see figure 4.5.

Then the various arguments are connected each other by the OR operator but, it is also possible to set up a tie by the AND operator, as shown in figure 4.6.
A further choice consists in using the linguistic expansions throughout hyponymy, hyperonym and synonym relations between terms. After preparing the query, clicking on run button the system creates a new window, where are presented the results: the set of dispositions that satisfy the criteria set up in the query prepared. The result encompass the following information: (see figure 4.7)

- disposition type
- formal partition (article and paragraph number)
- name argument
- value argument
- disposition text (on explicit user demand and extracted by a DOM parser that process the legal act in XML-DTD-NiR format)
The smallest information showed is the single disposition, therefore even if the user compose own query with condition on a single argument, the results, if not empty, will comprehend, at least, a single disposition with all its associated arguments. For example if the user choose

- disposition='obligation' and action='notification',

In the results page will be shown all the dispositions which meet the query conditions with all their arguments and not only the single argument action connected with the query prepared.

5 Correlated dispositions

In the legistic tradition some kind of provisions, in some kind of situations, are linked with references. Let s call them syntactic relations (explicit links).

For instance a derogation is usually linked to the derogated disposition, to make explicit this relation between the two provisions. The frequent lack of those explicit syntactic references makes harder legal order understanding.

The D-A meta data model can, to some extent, describe semantically relations and so allow us to capture some correlations between several dispositions types. Therefore derogative provisions, for instance, can be further asked for in a query, founded tanks to the explicit reference, if existing, or also tanks to common arguments contents and to the type of provisions involved. Let s call them semantic relations (conceptual or implicit links).

Semantic relations between dispositions have basically two foundations, logic and technical (legistics).

Basic legal concepts, to which several dispositions belong, are related according logic rules. Constituives and Regulatives class are mainly concerned with those relations (logic of norms).

Those also are detectable through arguments contents. At present three logical relations have been individuated and activated. First one the well known logic relations between deontic concepts are used as relations between the homonymous provisions, allowing for example, when looking for obligations of a particular subject, to find also implicit obligations that take the form of claims of his counterpart. From those correlations implicit rules can be inferred.

According to legistics rules and legislative drafting practices, legislator can and uses establish explicit relations between several kind of disposition (syntactic relation).

Regulatives - Violations: Violation normally references to a functionally related Regulative disposition.

Rules - Changes: a whatever kind selected Rule can be influenced and changed by Modifications and Derogations; in this case it is normally referenced.

As explained before, provisions that belong to the Changes class (Derogations and Modifications), can be marked up as Changes but also as Rules; this strategy allows
us to capture and find derogations and modifications provisions also tanks to their semantic features.

For instance a legistic relation lies regulative and violation type provisions: when they rule the same case in point, they have the same addressee and action arguments contents. In the Italian legistic practice regulative provisions and the connected violation dispositions are usually (but not systematically) also linked by a reference.

Related rules can be found activating syntactic or semantic-legistic relations.

Those two kinds of connections, semantic and syntactic, allows an expansion of the answer, adding, if requested, new related provisions and so giving the answer more advice completeness. The metaSearch through the correlation dispositions request allows the user to retrieve further dispositions, to be added to the result of a query, and allow to show the complete result as union of the two sets: the first one extracted from the first query and the second one made by the correlated dispositions. The aim is in general to turn as much as possible searching into advice, as a guide to the legislative texts consultation.

6 Queries Expansion : metaSearch

In the results page, according to the prepared query, it can appear a combo box in order to query the correlated dispositions, shown in figure 6.1
As we said, there are two correlation types:

- syntactic correlations
- semantic correlations

The syntactic correlation is based on the existence of an explicit normative reference between the involved dispositions, for example in the action argument of a violation disposition type has been set up a normative reference to an obligation disposition type.

\[(\text{violation, action}) \rightarrow (\text{obligation})\]

The syntactic correlation is already existing in the DTD-NiR formal structure, nevertheless, in the semantic mark up, it carries out through a normative reference term, a special mark up that indicates an explicit normative reference to the partition of one disposition.

Such reference is expressed also in the NiR-DTD formal profile, through a standard that comes from XPath syntax, however the formal profile does not distinguish between references normative and informative reference.

The semantic correlation can be seen as a dyadic or 2-place relation: it puts together a number of pairs (Disposition, Argument), greater than or equal to 1, so defined:

\[(D_x, a_x) \rightarrow (D_y, a_y)\]

where \(D_x\) is a \(x\) disposition type and \(a_x\) is an argument belonging to \(D_x\), \(D_y\) is a \(y\) disposition type and \(a_y\) is an argument belonging to \(D_y\). For example:

\[(\text{violation, action}) \rightarrow (\text{obligation, action})\]

In this example the correlation is made by one single pair, an other more complex example is the following:

\[(\text{obligation, addressee}) \rightarrow (\text{claim, counterpart})\]

\[(\text{obligation, counterpart}) \rightarrow (\text{claim, addressee})\]

\[(\text{obligation, action}) \rightarrow (\text{claim, action})\]

The pairs can be put in relation between themselves, that is they can make part of the correlation, only if the associated arguments have in common the keywords. In other words the arguments involved in the relation contain the same terms extracted from the controlled legal vocabulary.

The correlation is a symmetric relation and therefore this implies that if I build a query to extract the obligations dispositions type and then I request the correlated claims dispositions type, I must get the same result, starting from the claims and then requesting the correlated obligations.
If the correlation pairs are greater than one, we can classify the correlation also in
according to the following types:
* strong: if all the pairs satisfy the equivalence between the arguments
* weak: if at least a pair satisfies the equivalence between the arguments

The information retrieval algorithm used by metaSearch to extract the correlated
dispositions can be represented also from a mathematical point of view, given the set

\[ D_X = \{ \text{dispositions satisfy the initial query} \} \]

then the correlated dispositions retrieved correspond to the set \( D_C \):

\[ D_C = \{ D_X \cup D_Y \} \]

where

\[ D_Y = \{ D_j : (D_i, A_j) \} \]

\[ (D_i, A_i) \text{ where } D_i \in D_X \]

taking the correlation symmetric property into consideration.

A special case of semantic correlation concerns the semantic exceptions, in this
case the information retrieval rely on the multiple semantic mark up strategy. Let's see
an example:

article 10, paragraph 5: The right to obtain communication of the data in intelligible
form does not apply to personal data concerning third parties, unless breaking down
the processed data or eliminating certain items from the latter prevents the data
subject's personal data from being understandable.

This partition represents an exception to a right disposition type and so it will be
mark up also in the following way:

<table>
<thead>
<tr>
<th>partition</th>
<th>art10-par5</th>
</tr>
</thead>
<tbody>
<tr>
<td>disposition type</td>
<td>right</td>
</tr>
<tr>
<td>argument activity</td>
<td>communication</td>
</tr>
<tr>
<td>argument addressee</td>
<td>data subject</td>
</tr>
<tr>
<td>argument object</td>
<td>personal data</td>
</tr>
</tbody>
</table>

This will allow the user to retrieve the art10-par5 partition as exception of the
art7-par1 partition as right disposition type because the two dispositions are
correlated themselves for the arguments equivalence.
7 Queries Samples: metaSearch

Now we provide some useful query samples to emphasize the advantages of the semantic mark up from the information retrieval point. The first example, shown in figure 7.1,

![Fig. 7.1: obligations about the personal data assignments query sample](image)

concern the obligations about the assignment of personal data and it is built in metaSearch in the following way:

<table>
<thead>
<tr>
<th>disposition type</th>
<th>obligation</th>
</tr>
</thead>
<tbody>
<tr>
<td>argument action</td>
<td>assignment</td>
</tr>
<tr>
<td>argument object</td>
<td>personal data</td>
</tr>
</tbody>
</table>

The arguments involved are connected each other with the logical operator AND; the results, one single disposition of obligation type, are showed in Figure 7.2:
Then if we ask the system the violation disposition type syntactically correlated, in other words if we want to find out all the violations we'll face in the case I don't obey the obligations, we'll get the results depict in figure 7.3; the correlated violations are stresses with a different color and listed above the disposition associated:

![Fig. 7.3: violations syntactically correlated concerning obligations about the personal data assignments](image)

A second example concerns the data subject rights about communication, and is built in *metaSearch* as:

<table>
<thead>
<tr>
<th>disposition type</th>
<th>right</th>
</tr>
</thead>
<tbody>
<tr>
<td>argument <em>addresssee</em></td>
<td><em>data subject</em></td>
</tr>
<tr>
<td>argument <em>activity</em></td>
<td><em>communication</em></td>
</tr>
<tr>
<td>argument <em>object</em></td>
<td><em>personal data</em></td>
</tr>
</tbody>
</table>

The results, using the AND boolean operator between the three arguments, are presented in figure 7.4:
Starting from these result set the program allows the user to retrieve the violations semantic correlated. As in the previous example the semantic correlated violations are stresses with a different colour and listed above the disposition associated, see figure 7.5:

An other meaningful example concerns the data controller obligations; according to its definition, the data controller shall mean any natural or legal person, public
administration, body, association or other entity that is competent, also jointly with another data controller, to determine purposes and methods of the processing of personal data and the relevant means, including security matters, obligations regarding the data subject. That query is built in metaSearch as:

<table>
<thead>
<tr>
<th>disposition type</th>
<th>obligation</th>
</tr>
</thead>
<tbody>
<tr>
<td>argument address</td>
<td>data controller</td>
</tr>
<tr>
<td>argument countepart</td>
<td>data subject</td>
</tr>
</tbody>
</table>

The results, using the AND boolean operator between arguments, are presented in figure 7.6:

![Fig. 7.6: data controller obligations regards the data subject result](image)

Starting from these result set metaSearch allows the user to retrieve the claims semantic correlated. that is the implicit obligations, according to a deontic relation between obligation and claim. This type of correlations enable the user to catch different law maker's style: in other words an obligation can be expressed also as a claim of the counterpart. The results are shown in figure 7.7 and as usual the correlated claims are stressed with a different colour and listed above the disposition associated:

![Image with correlated claims](image)
A last example rely on the disposition class devised in D-A model and the ontological classes introduced in ontoPrivacy: it deal with the competences attribution to authorities. This examples is constructed in metaSearch in the following manner (see figure 7.8):

<table>
<thead>
<tr>
<th>disposition type</th>
<th>Attribution class</th>
</tr>
</thead>
<tbody>
<tr>
<td>argument address</td>
<td>Authorities as role introduced in ontoPrivacy</td>
</tr>
</tbody>
</table>
The results, see figure 7.9, consist in a rather big number of dispositions belonging to Attribution class dispositions:

8 Conclusions: Searching for Rules rather than Documents

Surface descriptions of a text results into defining documentary units, such as articles and paragraphs. Pragmatic and semantic contents description models add virtual documentary units of substantive character to information systems: legal provisions and their arguments contents.

The traditional search for legal texts or articles will be coupled with the conceptual search by normative provisions and their semantic contents. The system will be required to yield units (articles, paragraphs, etc.) including specific types of provision to be found in various laws at the same time.

This will allow easily identifying all the regulative contents provided for by all the laws regulating a given subject matter or case in point.

References

18. B. McLaughin: Java and XML OReilly 2001
20. S.Power: Practical Rdf OReilly 2003
22. E. Sachs: Getting started with Protegé frames Standford University 2006