Eliciting Concepts from the Brazilian Access Law Using a Combined Approach

Priscila Engiel
PUC-RIO
Rua Marquês de São Vicente 225, Gávea, Rio de Janeiro
(55 21) 3527-1001
pengiel@inf.puc-rio.br

Claudia Cappelli
UNIRIO
Av. Pasteur, 296 - Urca, Rio de Janeiro, 22290-240
(55 21) 2542-1919
claudia.cappelli@uniriotec.br

Julio Cesar Sampaio do Prado Leite
PUC-RIO
Rua Marquês de São Vicente 225, Gávea, Rio de Janeiro
(55 21) 3527-1001
www.inf.puc-rio.br/~julio

ABSTRACT
Lately, organizations have been subject to regulation promoting information transparency; one example of this is the Brazilian Information Access Law. This paper presents a novel way of performing requirements elicitation using both the law and a Non-Functional Requirements Patterns catalog as the information sources. Since organizations must follow the law, its information systems must also implement the law as requirements. Our process is guided by pattern matching, text mining and grounded analysis. We examine the special case of the Brazilian Access Law using our approach, which compares a previously encoded transparency knowledge base with the law.

Categories and Subject Descriptors
D2.1 [Requirements/Specifications]: Elicitation methods

General Terms
Your general terms must be any of the following 16 designated terms: Documentation, Design, Human Factors, Theory, Legal Aspects.

Keywords
Transparency, non-functional requirement, NFR framework, requirement pattern, text mining, knowledge extraction

1. INTRODUCTION
Organizations have been asked to provide transparency regarding their performance, management and outcomes. In the public context [1], the movement of open access can be seen as an important step to widen democracy and citizen participation in public matters. The demands for transparency are also borne by the need to satisfy regulations [2][3][4][5][6]. National and international organizations [8][9][10] seek to discuss the issue and create a knowledge network in this subject, also suggesting practices to achieve transparency.

SAC'14, March 24-28, 2014, Gyeongju, Korea.

http://dx.doi.org/10.1145/2554850.2555057

In 2011, the Brazilian Information Access Law [12] was promulgated. It regulates the public access to government information. The Law is a federal law applicable to federal, state, and county government. As such, the law is promoting a national policy to broaden the concept of transparency in Brazil and to prevent corruption. The main idea is to increase public participation and social control of government actions, by making possible that citizens do access public information, hoping for improvements in public management.

The law establishes, as a fundamental principle, that public information access is a rule and secrecy is an exception. To ensure the access, anchored on citizen's constitutional rights of, the Law defines the mechanisms, procedures, and deadlines for delivery of information as requested by citizens. The law also rules that public entities should disclose information proactively, mainly on the Web [20]. The law was first drafted by the executive branch within the CGU (Controladoria Geral da União)1 inspired by the work developed under UNESCO on Freedom of Information [18].

The main concern of the law is access, which contributes to transparency. Although allowing public access to information contributes to transparency, we understand transparency in a broader sense, whereas access to information is one of the components to its semantics. The work of Cappelli and Leite [11][13][14] aims to pin down the Transparency concept by building a transparency knowledge base described as a network of non-functional requirements (NFRs). According to them, Transparency is helped by the Accessibility, Usability, Informativeness, Understandability and Auditability NFRs. These NFRs are related by help contributions, a special relationship defined by the NFR Framework [17], and modeled as a Softgoal Interdependency Graph (SIG)., The knowledge base is further organized in NFR Patterns, following the proposal of Supakkul et al. [15].

The Transparency Patterns were developed based on [15], but with the addition of special patterns, one of them the GQO (Goal, Question, Operationalization) pattern, also known as Question pattern [17] that helps the refinement of each SIG leaf softgoal. The knowledge embodied in the GQO pattern supports the process of identifying good practices that operationalize the leaf NFRs of a SIG.

[1] "Office of the Comptroller General (CGU) is the agency of the Federal Government in charge of assisting the President of the Republic in matters which, within the Executive Branch, are related to defending public assets and enhancing management transparency through internal control activities, public audits, corrective and disciplinary measures, corruption prevention and combat, and coordinating ombudsman's activities." [10]

This paper aims to elicit Brazilian Access Law concepts with a combined approach, which uses: grounded analysis, text mining, and the NFR Patterns catalog. Central to this combination is the Transparency NFR Patterns which is used to: **a**) verify how the quality based reuse [19], provided by these patterns, could contribute to the interpretation and implementation of the law, and **b**) verify how the law contributes to achieve transparency.

The paper is organized as follows: Section 2 discusses related work. Section 3 presents the Non-Functional Requirements Patterns built for transparency. Section 4 presents how we did elicit concepts from the law, explaining the method we have used. Section 5 describes the comparison between the law elicited concepts and the catalogue. Section 6 summarizes our collaboration and point to future work.

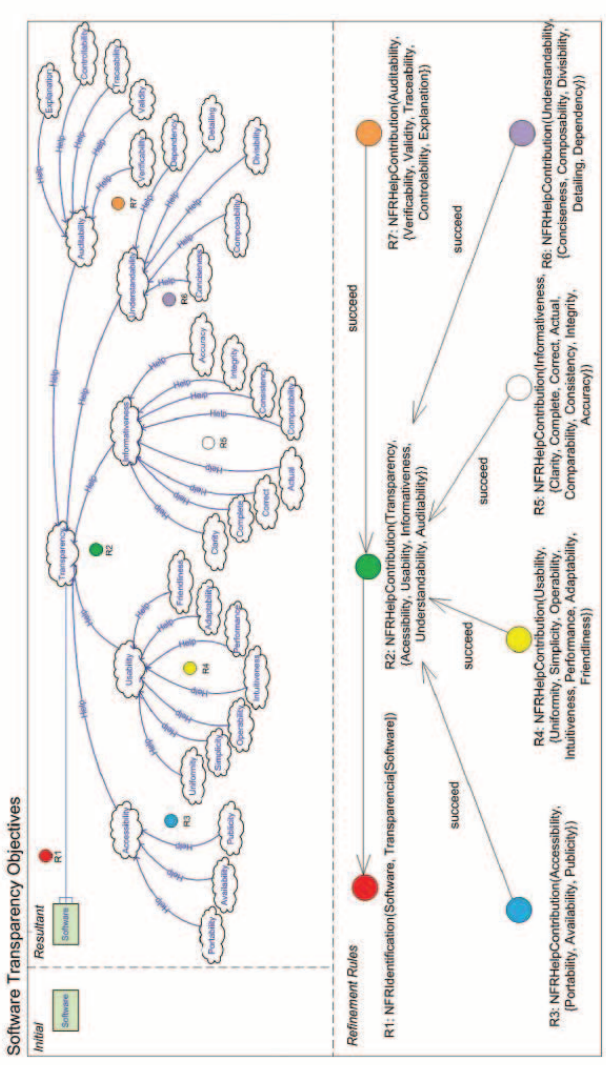

Figure 1 Software Transparency Objective Pattern [18]

2. BACKGROUND RELATED WORK

Motivated by the fact that Transparency is an important factor in improving democracy, the Requirements Engineering group at PUC-Rio has been working in eliciting the concept and modeling it as a NFR, using the ideas of the NFR Framework [26]. While Cappelli [13] studied ways of applying Transparency towards organizations, others in the group focused on Software Transparency [17] [14] [26]. Yet, another work, in partnership with UNIRIO, is developing a Transparency Maturity Model [27].

Cappelli [11] defines the concept of organizational transparency as a guide towards transparency adoption, providing to the organizations ways to implement transparency for their processes and information. Organization of this knowledge enables organizations to have the knowledge about how to integrate transparency in their policies, standards and procedures. Likewise, in software production, having the possibility of NFR knowledge reuse [19] will enable the production of transparent software.

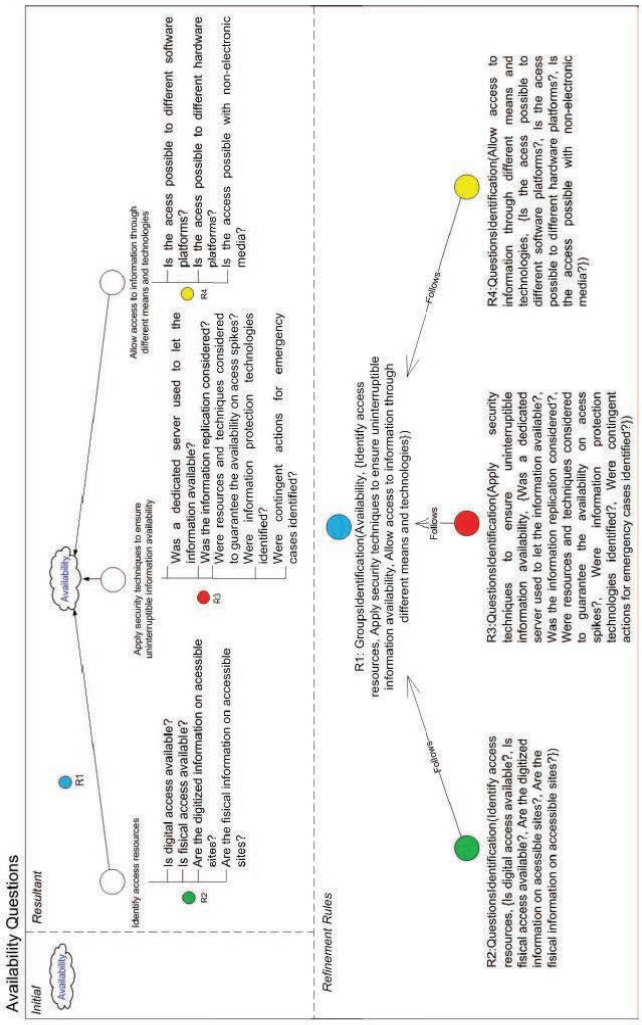

Figure 2 Availability Questions Requirements pattern [18]

Leite and Cappelli [14] summarize the main information sources used in the Transparency elicitation: "Four books were influential in our understanding of transparency. Holzner and Holzner (2006) provide an in-depth study from the social and historical perspectives on what they see as a movement to open government, in which transparency is key to achieving more open and democratic societies. Henriques (2006) examines different

constituents of transparency as a concept and frames them in the context of organizations, claiming that transparency will be essential for successful organizations. Lord (2006) provides arguments showing that increasing levels of transparency do not imply more democracy and peace as such insights are located at the limits of transparency. Fung et al. (2007) use the concept of target transparency as a way for organizations to reduce particular risks or performance problems through selective disclosure and does this by providing a careful analysis of the constituents of transparency". Two of these books were found by means of a systematic review process that found other information sources [15] from which knowledge on Transparency was elicited.

The Transparency Maturity Model [27], is a initiative towards establishing an evolutionary path for organizations to implement the transparency concept, achieving real gains from this initiative, both internally and externally. The model organizes the transparency characteristics in 5 levels or stages: opaque, disseminated, understood, trusted and participatory. Each maturity level/stage embodies a set of conditions that, when applied to the set of information that the organization discloses, will determine the maturity level of that organization. The set of conditions is a function of the transparency knowledge base (Transparency Patterns, as mentioned in Section 1).

3. REQUIREMENTS PATTERNS BUILT FOR TRANSPARENCY

The modeling of this knowledge took several years up to its actual representation as a collection of Transparency NFR Patterns [17]. Transparency is seen as a NFR because it is a quality issue; that is, it is orthogonal to the main functionality. Second, the characteristic is general, and as such, spreads to different parts of a given function. Third, it is not amenable to the typical measurement treatment as the one applied to functional characteristics; that is, we cannot say that something is or is not transparent. We will need to use a less objective judgment, like almost transparent, or transparent enough, and so on, and thus using the ideas of Herbert Simon on bounded rationality [36]. In the end, every RNF will be implemented as functions by "operationalizations", which are orientations about how the soft-goal can be implemented in the artifact that is studied.

The representation of Transparency as a collection of patterns is grounded on an objective pattern. Fig 1 shows the Transparency Softgoal Interdependency Graph (SIG) represented as an objective requirements pattern applied to software. Seven refinement rules are necessary to express the Transparency SIG. The first refinement rule, R1, relates the Transparency NFR to software. R2 defines the contributions between the five main transparency-related NFRs and the Transparency NFR (Acessibility, Usability, Informative, Understandability, and Auditability).). R3 to R7 define the contributions from twenty eight NFRs to the five main transparency-related NFRs. For example, R2 states that the Accessibility NFR helps the Transparency NFR. Moreover, R3 states that Portability, Availability and Publicity help the Accessibility NFR.

Fig 2 shows the GQO questions for the Availability NFR represented as a Question requirements pattern. The Groups Identification refinement rule, R1, defines three questions groups for the Availability NFR. Each Questions Identification refinement rule, R2, R3 and R4, lists the good practices questions pertinent to each group. For example, R1 states that "Apply security techniques to ensure uninterruptible information availability" is a question group of the Availability NFR. R3 lists five questions that help indentifying good practices, or "operationalizations", that answer to these questions.

The elicitation of the Transparency concept, as well as its modeling as an NFR Pattern were performed without any interaction with the Brazilian Information Access Law. As such the set of patterns, as a Transparency knowledge base, is independent of the law, and could be used as template to check if the law covers Transparency issues.

4. ELICITING CONCEPTS FROM THE LAW

In our work in the context of providing a transparency maturity model [27] it was mandatory to study the Information Access Law. To perform the elicitation of concepts from the text of the law, we have used text mining [28] [29] and grounded theory [30] [31]. Text mining was used because the strategy of elicitation was based on term frequency analysis and relationships among terms [22], following standard practices [22], [23] which equates higher frequency with relevance. The concepts were extracted automatically, in a similar way of automatic knowledge extraction of texts [25].

Grounded theory [30] [31] was used because we wanted to elicit knowledge from the law from the point of view of discourse analysis. The objective was not to create a theory, but we have performed a grounded analysis on the text of the law, as such we have used only the steps that guide the elicitation of concepts from the law, as such we have used: a) problem identification; b) sample construction and c) coding procedure.

In the first step, the problem defined was "How Brazilian Access Law is structured?" For the second step, sample construction, the text of the Brazilian Access Law (5,963 words) was used. It was not necessary to include other documents, because the idea of this part was to understand and extract knowledge from the law, without help of another document.

The third step was the coding procedure. First of all, open coding was performed, where subjects (concepts) were discovered. For this we have used text mining [28] [29]. As a mechanism, the software tool ATLAS.ti, which identifies the occurrence of each word that appears in the selected text, was used. To find the frequent subjects was necessary to perform pre-processing and remove the stop words (set of words considered irrelevant, like articles, prepositions, conjunctions). Using the tool we have found the most 15 frequent subjects in the law, as shown in Table 1. As previously stated, it is assumed [22] [23] that words with higher frequency are the words with higher relevance in the text.

WORDS	Total Count
information	87
access	70
information	68
law	64
public	53
term	30
entities	25
organs	25

authority	22
classification	21
secrecy	21
entity	20
caput	18
disclosure	18
organ	17

Table 1 - Most Frequent Terms

Another step was to perform the axial coding, which differentiates and refines categories. For this technique we used morphological normalization that consists in eliminating the morphological variations of a word, e.g. information and information's; law and legislation; access and accessibility. After this normalization the 15 most frequent categories in the law are shown in Table 2.

WORDS	Total Count
information	155
access	71
law	67
public	55
secrecy	48
entities	45
organs	42
term	39
classification	36
authority	32
regulation	23
administration	23
disclosure	20
procedures	19
disposed	17

Table 2 Normalized Terms

The most frequent terms were used to define the categories and relationships between these categories. We used the proposed relationships provide by ATLAS.ti and a concept map (diagram that shows relationships between concepts) was built (Figure 3).

The elicitation goal was to acquire the main concepts of the law, using a **combined** approach. In this approach once we have the conceptual map (Figure 3) we then verify it against the related NFR patterns, in this case the set of Transparency NFR patterns. The verification process used, as well, text mining to find pattern´s keywords in the text. As such, we are checking if the law covers key concepts of the Transparency knowledge base.

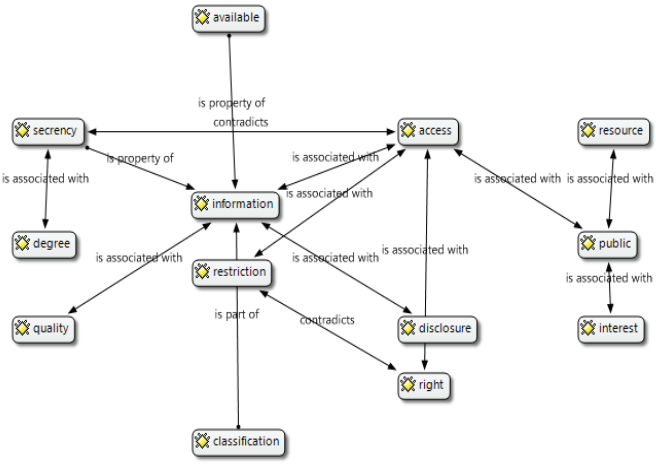

Figure 3 Access Law Conceptual Map

5. OBSERVATIONS FROM THE ELICITATION PROCESS

Out of the 33 transparency catalogue characteristics (*accessibilility*: portability, availability, publicity; *auditabilility*: validity, controllability, verifiability, traceability, accountability; *informativeness*: consistency, integrity, accuracy, completeness, clarity, comparability, correctness, current; *understandability*: conciseness, composability, decomposability, extensibility, dependability; *usability*: uniformity, simplicity, operability, performability, adaptability, user-friendliness, intuitiveness.), only 10 were found in the law: "accessibility", "availability", "publicity", "decomposability (detailed)", "informativeness", "clarity", "integrity", "correctness", "current" and "completeness". And of these 10 only 2, "availability" and "accessibility" appears in the conceptual map (Fig. 4). However, we find "disclosure" in the map, related to information, which can be interpreted as informativeness. We understand that the combined approach, by providing a certain degree of redundancy does help the construction of a rich mental map, including those characteristics that was cited but not included in the drawing.

Although not all characteristic of accessibility were present in law there are other characteristics present in law, since it refers to the characteristics of "informativess", like "clarity", "integrity", "correctness", "current", "completeness", and "understandability", and also mentions the characteristic of detailed information ("decomposability"). Analyzing the definition of "accessibility" in the knowledge base (NFR patterns), the law does not fit fully in that capacity because it does not mention the quality of "portability" as the Objective Pattern (Fig 1) does. The concept of "access" in the conceptual map (Fig 4) is concerned with disclosure information to the public, access in this case is the opposite of secrecy and restriction, and the definitions of "access" in the transparency NFR Pattern catalogue is broader: it is concerned of how a user can access the information in any place with or without technology at any time, as well as how it is publicized.

Another observation is that not all characteristics that appear in the law have a definition. Only "availability" and "integrity" are defined in law, and the definitions are semantically different of the definitions of the catalogue as presented in Table 3. We observe

that Transparency pattern catalog definition is more generic and applied in diverse situations, and the law definition is more specific.

Characteristic	Law definition	Catalogue definition
availability	Information quality can be know and used by equipament, people and autorizated systems.	Capacity to be accessed whenneeded.
integrity	Information quality that is not modified, including as to origin, transit e destination	Capacity to be a whole, in sense not to miss any part.

Table 3 - Characteristics definitions

Matching characteristics from the catalogue with those of the law enables the checking of GQO questions (Fig 2) for matched characteristics. So, it is possible to check if the law mentions operationalizations for a given characteristic. Figure 2 portrays the questions for "availability", which will lead to different sort of operations as to answer the questions. For instance, one of the questions is : "Are the digitalized information on accessible sites?", and the law says that it was mandatory to disclosure information's in sites of the World Wide Web, GQO questions pattern also says that is necessary to ensure access of information in different media and technologies, the law cited availability of information in print, electronic or any other means of universal access. The pattern question about uninterrupted availability of information is not mentioned in the law.

For the characteristic of "publicity" only the GQO about access restriction is mentioned in the law, the other GQOs about detail information's and techniques multimedia for information dissemination are not mentioned in the law. The law does not cite the characteristic of "portability", but in its text it cites the use of open formats, which is one of GQO questions for the portability questions pattern.

6. CONCLUSION

In this work we present a novel way of performing elicitation. We use a common method to extract the knowledge from the law using grounded theory and text mining, but add the reuse of a previously encoded knowledge base [19] as a way to to verify the elicitation. The knowledge base concerns quality concepts and is modeled using a NFR patterns structure. The approach was performed in a real case, the Brazilian Information Access Law, resulting in a deeper understanding of the law.

There are other techniques to elicit requirements concepts from laws. For instance, [35] proposes a framework for modeling law in software system design. It provides a core set of concepts to enable exploring and selecting alternatives in a variability space defined by the law, they elicit the concepts using the sentence oriented interpretation and a graphical notation to visualize the models generated by these concepts. The difference between our work and this work [35], is that, in the case of the Access Law, the concepts are mostly non-functional requirements. It is also the case that we have used a coarse interpretation whereas a sentence by sentence is fine grain elicitation. Our strategy, more abstract, is based on discourse analysis, to elicit the quality related concepts present in the law. These concepts are then checked with an already available body of knowledge (Transparency patterns) thus empowering a quality reuse strategy [19].

The Brazilian Information Access Law focus mainly on just one aspect of transparency, as its name implies, access, however compared to the Transparency Patterns, it is observed that this characteristic is not fully implemented, since the characteristic for portability is not cited in the text law. One the other hand, the law lists other characteristics that are important in achieving transparency, but which are not directly related to access, like clarity, integrity, correctness, current , completeness. The law has a focus in providing and obtaining information, but fails to mention how to structure this information's in a way that everyone can use, understand, and check its contents.

This analysis shows that using the combined approach of text mining and grounded theory and comparing the result with NFR Patterns makes it possible the discovery of important concepts not in the law. We also observed the lack of definition to important characteristics to achieve access, like, for instance, availability that is cited, but not defined in the law.

In our vision of a Transparency Maturity Model [27] we consider obeying the Information Access Law a first step an organization must follow towards implementing Transparency. We are following the implementation of the Law and updating our Transparency Requirements Patterns with different processes possibilities to help organizations comply with the Access Law.

Future work will apply this approach to other laws involving quality characteristics. We also want to run this same experiment with other actors in order to compare the results. Another future work is to delve in mechanisms that will focus on the characteristic of *understandability*, since it is central to the quest of informing citizens.

REFERENCES

[1] Tapscott D., Ticoll D. (2003) "The Naked Corporation: How the Age of Transparency Will Revolutionize Business", Free Press, 1 ed.

[2] BASEL - Basel Committee on Banking Supervision - http://www.bis.org/. Access in 22/08/2010.

[3] EITI – Extractive Industries Transparency Initiative - http://eitransparency.org/. Acessado em 10/02/2013.

[4] OECD - Organization for Economic Cooperation and Development http://www.oecd.org. Access in 10/02/2013.

[5] Law 131 - https://www.planalto.gov.br/ccivil_03/Leis/LCP/Lcp131.htm Access in 10/02/2013.

[6] TCC – Transparência, Consciência e Cidadania - http://www.tcc-brasil.org.br/. 1996. Access in 10/02/2013.

[7] Fung, A.; Graham, M.; Weil D.(2007)" Full Disclosure: the Perils and Promise of Transparency". Cambridge University Press, Londres, UK

[8] TB - Transparência Brasil - http://www.transparencia.org.br/index.html. 2000. Acessado em 10/02/2013.

[9] OGP – Open Government Partnership. (http://www.opengovpartnership.org/). Access 10/02/2013.

[10] CGU – http://www.cgu.gov.br/english/default.asp Corregedoria Geral da União -http://www.cgu.gov.br/CGU/. Access in 04/28/2013.

[11] Cappelli C., Leite J.C.S.P., Oliveira A.P.A. (2010): Exploring Business Process Transparency Concepts. RE 2007: 389-390

[12] Law 12.527 - http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm Access in 01/28/2013.

[13] Cappelli, C. (2009) "Uma Abordagem para Transparência em Processos Organizacionais Utilizando Aspectos". Rio de Janeiro, 2009. 328 p. Tese de Doutorado – Departamento de Informática, Pontificia Universidade Católica do Rio de Janeiro.

[14] Leite, J. C. S. P.; Cappelli, C.(2010) . Software Transparency. Business & Information Systems Engineering,Springer, p. 127-139

[15] S. Supakkul, T. Hill, L. Chung, T. T. Tun and J.C.S.P. Leite (2010): "An NFR Pattern Approach to Dealing with NFRs," Requirements Engineering, IEEE International Conference on, pp. 179-188, ISBN: 978-0-7695-4162-4.

[16] Chung L. , Nixon B. A., Yu E. , and Mylopoulos J. (2000): *Non-Functional Requirements in Software Engineering*. Kluwer Academic Publishers

[17] Cunha, H.; Sampaio do Prado Leite, J.C.; Duboc, L.; Werneck, V., "The challenges of representing transparency as patterns," Requirements Patterns (RePa), 2013 IEEE Third International Workshop on , vol., no., pp.25,30, 15-15 July 2013

[18] Mendel, T. (2008) "Freedom of Information: A Comparative Legal Survey", UNESCO: Paris, 2008

[19] Leite, J.; Yu, Y.; Liu, L.; Yu, E. & Mylopoulos, J. (2005): Quality-Based Software Reuse. In: Advanced Information Systems Engineering: 17th International Conference, CAiSE 2005, 2005, Porto, Portugal. Advanced Information Systems Engineering: 17th International Conference, CAiSE 2005, Porto, Portugal, June 13-17, 2005

[20] Controladoria-Geral da União – CGU, Acesso à Informação Pública: Uma introdução à Lei nº 12.527, de 18 de novembro de 2011, Brasília 2011.

[21] Harrison, T. M., Guerrero, S., Burke, G.B., Cook, M., Cresswell, A.,Helbig, N., Hrdinová, J, and Pardo, T. (2011)"Open Governmentand E-Government: DemocraticChallengesfrom a PublicValue Perspective", In: Proceedingsofthe 12th AnnualInternational Digital Government Research Conference. pp. 1-9.

[22] His I., Potts C., Moore M.M., (2003) Ontological Excavation: Unearthing the core concepts of the application. WCRE, page 345-352. IEEE Computer Society

[23] Krallinger, M., Erhardt, R. A. A. and Valencia, A., 2005. Text-Mining Approach in Molecular Biology and Biomedicine. Drug Discovery Today, Vol. 10, No. 6, pp. 439-445.

[24] Sayão, M. (2007) "Verificação e Validação em Requisitos: Processamento da Linguagem Natural e Agentes". Tese de Doutorado, Universidade Pontífica Católica do Rio de janeiro, PUC-RIO,

[25] Sebastiani, F., 2002 Machine Learning in Automated Text Categorization. ACM Computing Surveys (CSUR), vol 34, Issue 1, pages 1-47

[26]] Leite, J.C.S.P.(2009) Software Transparency, Dagstuhl Seminar Proceedings 08412.Perspectives Workshop: Science of Design: High-Impact Requirements for Software-Intensive Systems http://drops.dagstuhl.de/opus/volltexte/2009/1929

[27] Cappelli C., LeiteJ.C.S.P., Tanaka,A.K., Santos G., Araujo R.M., Nunes V., Engiel P., (2013) "Um modelo de estágios para avaliar transparência organizacional" Revista do Tribunal de Contas do Município do Rio de Janeiro, abril.

[28] Hearst, M. A. (1997) Text data mining: Issues, techniques, and the relationship to information access. Presentation notes for UW/MS workshop on data mining, July 1997

[29] Feldman, R. & Dagan, I. (1995) Knowledge discovery in textual databases (KDT). In proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, August 20-21, AAAI Press, 112-117

[30] Martin P.Y., Turner B.A.,(1986): "Grounded Theory and Organizational Research," *The Journal of Applied Behavioral Science*, vol. 22, no. 2 pp 141..

[31] Glaser, B.G. and Strauss, A.L., (1967) The discovery of grounded theory: strategies for qualitative research. Chicago.: Aldine

[32] Holzner B., Holzner L., (2006) "Transparency in global change: the vanguard of the open society". University of Pittsburgh Press, Pittsburgh

[33] Lord K.M. (2006) The perils and promise of global transparency. State University of New York Press, New York.

[34] Henriques A. (2006) Corporate truth the limits to transparency. Earthscan, London

[35] Ingolfo S., Siena A., Perini A., Susi A., Mylopoulos J., "Modeling Laws with Nomos 2", 6th International Workshop on RE and LAW: Special Track on Convergent Challenges, Rio de Janeiro. July 2013.

[36] Simon H A (1969). "The Sciences of the Artificial (First Edition), MIT Press.