Identifying Objective True/False from Subjective Yes/No Semantic based on OWA and CWA
Duan Yucong, Christophe Cruz, Christophe Nicolle

To cite this version:
Duan Yucong, Christophe Cruz, Christophe Nicolle. Identifying Objective True/False from Subjective Yes/No Semantic based on OWA and CWA. Journal of Computers (JCP), Academy Publisher, 2013, Vol 8 (No 7 (2013)), pp.1847-1852. <10.4304/jcp.8.7.1847-1852>. <hal-00927476>
Identifying Objective True/False from Subjective Yes/No Semantic based on OWA and CWA

Yucong Duan
Le2i, CNRS, University of Bourgogne, Dijon 21000, France
Email: duanyucong@hotmail.com

Christophe Cruz and Christophe Nicolle
Le2i, CNRS, University of Bourgogne, Dijon 21000, France
Email: {christophe.cruz, cnicolle}@u-bourgogne.fr

Abstract—Distinguishing objective(OBJ) semantic from subjective(SUBJ) semantic is essential to information processing of either explicit design formalization or implicit natural language(NL) communication of projects’ implementation. We summarized our past experience of solutions for semantic knowledge management projects. An adopted hypothesis is that among the tremendous and rapid increasing information, there is a relative stable core which can map/relate to every specific piece of information uniformly starting from the discussion on existence and conceptualization. The mapping will result in an expansion formalization based on open world assumption (OWA) and closed world assumption (CWA). This formalization can be widely applied for guiding semantic formalization and validation. We show initial applications on transformations from SUBJ to OBJ, etc.

Index Terms—Semantic, Knowledge, Cognitive, Design, Formalization, Conceptualization

I. INTRODUCTION

In the past, we have participated in a project on knowledge management for building objects identification with semantic[1, 2, 3,4]. It originally involves three general steps:

(a1) The knowledge is initially provided by refinement of the information provided by various stakeholders. If we define related information as every natural language expression by these stakeholders have the potential to be useful knowledge. The search space for potential knowledge is almost as big as the information space which is composed with unlimited natural language expressions.

(b1) And then the empirical knowledge is expressed as structured semantic rules sets corresponding to the internal relationships among them.

(c1) Rules are organized empirically for different processing sequences and reasoned with data instances to identify building objects.

Disadvantages of relying on empirical approaches:

- The semantics of the knowledge rules are implicitly managed in the minds. The management could be vague instead of clear.
- There are potential semantic gaps, overlaps and inconsistency among the expressed rules. This situation will be more obvious with the increasing of the amount of the knowledge rules.
- The reorganization of the rules for a new application will be very difficult as the semantic relationships among the rules are implicit.
- There is not a clear evaluation of what kind of targets can be achieved or not for the rules.

The rest of this paper is organized as follows. First we present the background hypotheses on semantics expansion and multiple semantics which are used for subsequent analysis. Secondly we analyze on the related semantics phenomena. Thirdly we present the strategy corresponding to the project analysis and hypotheses. After that we show some case studies of applying the proposed approach. In the end we summarize this work in progress with extensions on future work.

II. BACKGROUND HYPOTHESES

A. Hypotheses on Semantic Expansion of Natural Language and Web Semantic

In general information technology leads to an unprecedented increase of information share and exchange. Our hypothesis: We distinguish information expression from the view of human side as two parts of denotation notations and intended semantic.

(a) Static. Underlying the appeared huge amount of expressions, there is a less amount of semantic. The semantic can be further reduced to a less amount of essential semantic.

(b) Dynamic. The increase of the amount of expressions exceeds that of the semantic. This will pose the possible problem that there is the tendency that intended semantic will be increasingly buried unintentionally by a bigger amount of notations.
B. Hypotheses on Multiple Semantics with Single Expression

Hypotheses on multiple semantics: Taken a single expression, when viewed from different individual views and with different background information, usually multiple semantics can be identified. One common situation is that the information contained in the expression is incomplete or dependent on the supplement information which implicitly relies on the minds of specific individuals, e.g., the concepts of “left vs. right”, “behind vs. front”, etc. Another situation is when the expression is enquired with “whether the semantic represent all situations or not”. A general case is that usually a concept could be considered at instance vs. type level. We extend our discussion on both situations in subsequent paragraphs.

There may also be several semantics during processes of introduction of concepts by human. These semantics could be the contents which are either created through a conceptualization by human or created through other operations, e.g., abstraction and even by mistakes.

Hypotheses on processing: any absent or omitted semantic might need to be identified if they have a positive/negative impact on the fulfilling of a project. If they do not have an essential impact on the project, the fact has to be validated as well.

Then we need to cope with the vague, absent, mistakes of semantic which might not covered by the DL facility. But we do not want to process too detailed concrete project requirement specification specifically. We limit our ambition to reaching an on the fly approach which can be provide guidelines to detect and validate potential problems at both abstract semantic level and concrete semantic level.

III. Analysis of Semantic Phenomena

We meet quite a lot theoretical and engineering challenges which we identified as originating in natural language (NL) semantic processing. We list the challenges as follows:

(a2) Conceptualization level. We need an approach which differs from the empirical approach. We start from the hypothesis that among the tremendous and rapid increasing information, there is a relative stable core which can map/relate to every specific piece of information uniformly starting from the discussion on existence and conceptualization. Existence limits the basis of capable discussion. All other things including meaningful is only a consequence hence after. Conceptualization represents a cognitive process from observation to forming concept and hence after. Here we focus on the conceptualization from existence to forming concepts. Corresponding to (a1), stakeholders and even a single one could be unconscious [6] with the semantic of the NL expression which is expressed by them. Different conceptualization need to be unified or validated as a base for further semantic utilization. We could not unify the conceptualization after we receive an expression from stakeholders. Only validation approach could be explored.

(b2) Semantic content level. For (b1), we identified that the transformation from NL semantic into machine acceptable semantic involves classification (CLA) as implementation (IMP) vs. description (DES) purposes and order (ORD) these rules according to their sequential relationships. This actually demands a guide for creation knowledge rules which represent IMP rules while avoiding providing DES rules for machine to automatically implement.

(c2) Application level. For (c1), we identified that rules need to be organized according to different processing targets and optimized based on well engineering tradeoff. A unified expression for organization of application purposes and rules will be beneficial for easing the organization of the rules for application purposes.

IV. Solution Strategy

Firstly we present the solution strategy corresponding to the project analysis. After that a retrospect on the strategy is presented corresponding the solving of the problems of hypotheses.

A. Strategy Corresponding to Project Analysis

CWA and OWA are viewed by us as involving in any information statement with decisions on Y/N or T/F. While for most expressions, they are implicit or even get lost. We choose them as the basis for semantics analysis. An attempt to explain them as concepts could be easily frustrated usually since that most definitions [13] will be limited to the cycle of concept at conceptual level [6]. We argue that an understanding of CWA vs. OWA semantic is much more complex than it appears, e.g., CWA and OWA are exclusive to each other at the existence level while pure discussion will not forbid any expressions which compose both at the same time. The context needs to be clarified carefully to achieve formal semantics of them. To avoid to be trapped into cycles of explaining a concept using other concept at the conceptual level [6] when you inquiry about the ultimate understanding in your mind, we argue that it actually demands a conceptualization experience which start from observation and confirmation of existence towards forming concepts. A conceptualization of CWA vs. OWA can be found in [6].

The solution strategy corresponding to the problem analysis is proposed based on an analysis of CWA vs. OWA as follows:

(a3) Based on a dualism/CWA of human and material, we propose a conceptual validation approach for concepts. The dualism provides a basic complete category of existence of either human or material. Firstly we choose a concept and then we validate the proposed semantic as from the view of an individual human. A condition which has to be met is that every related semantic which is used for explaining the target semantic has to be traced back to the level of existence of the dualism.

For knowledge expression, we define an information as SUBJ as that the same expression could have different versions semantic interpretation. Similarly we define an information as OBJ as that the same expression will have
a unified interpretation by all. From the CWA view of this dualism, OBJ information can be identified as complete/CWA or independent while SUBJ information can be identified as incomplete/OWA since that the human side semantic/explanation is implicit and absent from the explicit expression [6]. A retrospective refinement of this explanation is that OBJ information can be used to deny any information which is different based on CWA while SUBJ information cannot since it is related to OWA.

A decision on SUBJ vs. OBJ can be made based on the following criteria when the individual are changed: if the semantic is the same after the change, the concept is classified as OBJ. If the semantic is different, it is classified as SUBJ. This is helpful for validation against the proposition of whether the proposed expression is SUBJ/OBJ or not. Then a remedy can be designed accordingly by explicitly adding the implicit human side semantic.

For a control operation of projects, SUBJ decision results in Yes/No(Y/N) and OBJ decision results in True/False(T/F). A throughout understanding of the terms of Y/N and T/F also demands a conceptualization experience which is detailed in [6]. If all these Y/N and T/F are identified for a control flow, a Y/N to T/F flow [10, 11, 12] can be achieved. It represents the necessary human SUBJ interactions which correspond to Y/N and feasible automation of a machine which corresponds to T/F. There is always a gapless “Y/N→T/F” flow which maps to an implementation of a project. The task of this step is to identify or supplement it. This approach is compatible with the common understanding of SUBJ vs. OBJ of which SUBJ could represent both results of Y/N and T/F. For this approach, it is a phenomenon of replacing T/F as T/F or assignment of tasks which are feasible for automatic processing to human.

(b3) After (a3), we progress to the validation of the organizations of the original Y/N → T/F flows against the criteria of IMP vs. DES. We propose some direct modes which support the optimization of the Y/N→T/F for simplification of the expressions as follows: CWA + CWA → CWA which can be mapped to T/F + T/F → T/F. It means that several connective independent information processing segments can be reduced to one single expression without influence the effect of this analysis purpose. Similarly we propose OWA + OWA → OWA which can be mapped to Y/N + Y/N → Y/N. We also propose some validation modes to detect the counterexamples as follows: CWA + CWA !→ OWA fits the situation of IMP by a machine. Also a flow containing OWA like “→CWA→…OWA…→ CWA” could not be IMP without supplemental transformations of “OWA→CWA” explicitly.

(c3) For (c2), we will apply the semantics semantic formalization approach to formalize requirement expressions of processing targets. This will result in a requirement expression in the same form of the organization and usages of the rules at the abstract level of “Y/N→T/F” flows guided with CWA vs. OWA validation: “Y/N→T/F” | CWA vs. OWA. This will facilitate the process of organizing rules to fulfill those targets.

B. Processing Corresponding to Hypotheses

For achieving the simplification of semantic (1), we propose a technique to map the semantic to an abstract level of <<Y/N, T/F>, transition> completely.

(a) For pure DES situations. Pure DES situations refer to the situations where expressions can not be automatically executed by a machine as intentioned in general. We can apply the reductions from transitional SUBJ→SUBJ or Y/N→Y/N to SUBJ or Y/N: “SUBJ→SUBJ”→ SUBJ or “Y/N→Y/N”→ Y/N.

Explanation: they are reduced since that only one SUBJ of Y/N must be reached if the sequential process has to proceed.

(b) For pure IMP situations. Pure DES situations refer to the situations where expressions can be automatically executed by a machine as intentioned in general. We can apply the reductions from transitional OBJ→OBJ or T/F→T/F to OBJ or T/F: “OBJ→OBJ”→ OBJ or “T/F→T/F”→ T/F.

Explanation: they are reduced since that all of the T/F process can be processed inside a black box by a machine for human as if there is only one OBJ of T/F.

(c) For mixed expressions. The mixed expressions of DES and IMP expressions need to be further segmented before being processed as 1) or 2).Then all the situations are guaranteed to be able to be processed by 1) and 2).

After applying the procedures, a sequence composing repeatedly “SUBJ→OBJ” can be achieved. The sequence is the expected final semantic representation. In comparison with the original representation, the semantic overlap can be reduced.

If the final sequence is some isolated sequences instead of an integrated sequence, there are the possibilities that there are some gaps or vagueness in the original expression. Then some remedy work could be designed and adopted to fill the gaps.

V. CASE STUDIES

The discussion of the cases is extended at both language/logic expression mechanisms and a concrete project related contents.

A. On semantics of Logic Connectives and Connected Concepts

For description logic (DL), the semantic of logic connectives can be restricted in a formal manner. We observed that there are more than one semantic with the logic connectives and there may be even more semantic with the composition of logic connectives. Some of these semantic are not restricted by DL. And the meaning of them relies on the definitions of human. For these situations, there are possibilities that the expected semantic will be omitted or wrongly supplied by human. All of them could result in a transitional controversial
decision. For the semantic phenomena of classical logic connectives, when CWA vs. OWA are queried for the backgrounds of specific logic connective in an expression, there will be several semantics which need to be refined. For logic connectives, we have extended a discussion which is based on revelation of the conceptualization of logic connectives in [13]. For the semantics of concepts, we will show how the proposed approach will apply.

B. Application for Suppliment DL Expressions

(d) For logic connectives. We supply more semantic by adding explicit CWA/OWA as backgrounds. Among these semantic, there might be some semantic which are omitted or neglected. Also some mistakes of semantic usages/descriptions might be identified during the process from implicit to explicit of the CWA/OWA backgrounds.

(e) For the concepts of entities and relationships. Based on throughout CWA/OWA backgrounds analysis, we propose to firstly identify the vague semantic as Y/N vs. T/F through explicitly identifying the backgrounds of CWA vs. OWA. The involved formulas include:

- For relationship concepts. The OWA based dependent/incomplete conceptualizations result in “left vs. right”, “vertical vs. horizontal”, etc. These semantic are not independent as that they actually rely on the reference to specific human being. These complete semantic are SUBJ instead of OBJ. We denote them as “leftSUBJ vs. rightSUBJ”, “verticalSUBJ vs. horizontalSUBJ”, etc.

The CWA based independent conceptualizations result in Y/N→T/F, e.g., “horizontalOBJ vs. verticalOBJ”. This process is specifically finished by fixing the achieved semantic of “verticalSUBJ vs. horizontalSUBJ” first, then change the specific human being concept to see whether the semantic of “verticalSUBJ vs. horizontalSUBJ” are still validated. This validation criterion could also be adopted for distinguishing among OBJ vs. SUBJ. Subsequent negation on the concepts could be implemented after a CWA is assumed. The semantic of “leftSUBJ vs. rightSUBJ” can not be transferred from SUBJ to OBJ as that after changing the individual human being concept, we find out that there is not enough information to maintain the original semantic. This is because of that a NL statement of “on the left of person(A)’s face/back direction”. We also call this kind of situation as unconscious (vs. conscious) [6].

- For entity concepts. We believe that a semantic clarification based on the relationships’ clarification will help to achieve a similar goal for entities. The missing link is the conceptualization process from entity concepts to relationship concepts.

C. Building Objects Semantics Reconstruction Related Cases

We would like to share some of the cases which we have identified in the past during the process of semantically reconstruction of building objects. Situations of possible confusions are classified and described as follows.

We do not consider the special cognitive situations as that: a cube cannot be considered geometrically as “vertical/horizontal”, etc. We apply the backgrounds of CWA vs. OWA at the abstract but complete description level of SUBJ vs. OBJ and Y/N vs. T/F. The result will help to identify the incomplete semantics which could be taken as complete semantics.

Cases of “left/right”, “front/back”, etc. When speakers describe spatial objects with concepts like “left/right”, “front/back”, they will be quite clear with the semantics which they want to transfer with their expressions which contain these terms. In our cases, this information could be identified as a piece of knowledge which will be utilized for a machine to process automatically. Then the problem is that how to guarantee the OBJ of the information. Actually audiences of the expression of the information will explain the information with semantics which might be different from the intention of the speakers. This is because of that the speakers do not recognize that the semantics which they want to transfer are only partially explicitly expressed with their expressions. The other part of the semantics implicitly relies on themselves, e.g., their location and facing directions for this case. The absence of this information will make the intended independent information actually dependent. The explanation can be mapped to SUBJ→Y/N instead of OBJ→T/F. There is the need for information model transformation(MT) of MT(SUBJ→OBJ) by way of MT(Y/N→T/F).

Identification method: these terms will not reach the same semantic when the users are switched. So their semantic are not complete. This can be modeled and proved with the follows:

We model the original situation with an OWA background.

(a) Expectation

The transferred semantics are independent/complete or OBJ,

\[
\text{User(semantic(term(X)))=CWA} \\
\text{Instances: UserA(semantic(term(X)))} = \text{UserB(semantic(term(X)))}
\]

(b) Observation

They might be different.

\[
\text{User(semantic(term(X)))=OWA} \\
\text{Instances: UserA(semantic(term(X))) != UserB(semantic(term(X)))}
\]

Analysis:

We model the situation with a CWA background where human side semantics are also modeled explicitly.

(f) Expectation
Semantic(userA) + semantic(term(X)) = CWA
Instances: Semantic(userA) + semantic(term(X)) = semantic(userB) + semantic(term(X))

(g) Observation
Semantic(userX) + semantic(term(X)) = OWA
Or Semantic(userX) + semantic(term(X)) != OWA
* These two expressions are not equal if their validation scopes are considered.
Instances: Semantic(userA) + semantic(term(X)) != semantic(userB) + semantic(term(X))

(h) Additional conditions
Since that the expressions of semantic(term(X)) can be observed and not changeable then it can be assumed as CWA: semantic(term(X)) = CWA.

(i) Conclusion
By applying semantic(term(X)) = CWA to “Semantic(userA) + semantic(term(X)) = OWA” using the mode of “CWA + OWA = OWA” or “CWA + CWA != OWA”. It can be inferred that: Semantic(userX) != CWA or Semantic(userX) = OWA.

(j) The remedy
While specific Semantic(userX) are CWA at instance level, there is the need to transform from “Semantic(userX) = OWA” to “Semantic(userX) = CWA”. The incompleteness of OWAs could come from the situations that the human side implicit semantic information is omitted/neglected. This implicit information need to be considered if applicable. The transformation of MT(implicit → explicit). What is done can be explained as that implicit human side semantics are explicitly added to the expected transfer contents which turn the original incomplete semantics of the concepts for the projects complete: human side semantic + OWA+semantic(“termX”) → CWAs. In the case of “left/right”, the information on human side is: the positions of the individual which are demanded at instance level while the expression is transferred.

* When we talk with NL terms which might be at the instance vs. type level. We need to distinguish them. To simplify the expression, we sometimes explicitly restrict the terms’ semantics to the instance level by default.

In the project [10], we further identified the difference between “left vs. right” and “vertical vs. horizontal”. After the conceptualization, when the semantics of “left vs. right” and “vertical vs. horizontal” need to be transferred, both of them demand the supplement of the appearance of a human. After the confirmation or the explicit supplement of “Yes or No” is done by human, the newly composed semantic of “vertical vs. horizontal” can be transferred. This process is a case of semantics MT of SUBJ→OBJ. Semantic of “left vs. right” cannot be transferred even with the supplement of the semantics of “Yes or No” on left and right. We identified that the appearance of human is not enough. The facing direction of the human is also needed. This difference among semantics of concepts will be easily omitted if a similar conceptualization process is not performed.

VI. SUMMARY

In this draft, we describe the methodology which is employed for semantics clarification for knowledge management purposes. Through a discussion starting from existence to conceptualization in an evolving manner, an approach is proposed towards the seemingly unlimited natural language expressions of information. CWA and OWA are employed to map semantics to the level of Y/N and T/F in the manner as flows. The approach can be utilized in an on the fly manner.

The challenges for practice are listed as follows: There are several levels of semantics for a single expression. Usually we only analyze one or some of them. Some of them are not analyzed deliberately or the understanding of that semantic by itself demands a conceptualization experience other than any explanations. This will still leave space for possible misunderstandings of an expression even after some semantics of the expression are formalized. Although we observed that the type/class level semantics discussions are independent from the instance/object level semantics discussion from a constructive conceptualization background, people will get confused when they omit the existence of the conceptualization processes. So much more efforts than it appears to demand will be demanded to implement this approach for a project.

In the future, plan to introduce some enlightening mathematics findings like four color theory [8], etc. as guidelines for the conceptualization process to reduce the chances of making mistakes by human during the conceptualization processes. We also want to create a quality model to support, guide and evaluate the tradeoff and decisions during the semantics process for a specific project. It will be firstly translated/formalized with the proposed approach at the abstract semantic level. Then it will be directly available for aiding in achieving decisions on: What can be done? How to do? What to do? How to adjust/modify? Have these goals been reached? A glance of this work extended on the existence of multiple semantics can be found in [14, 15,16].

REFERENCES
Yucong Duan is currently a Post-doc fellow in Le2i, CNRS, University of Bourgogne, France, since June 2009. He received a PhD in software engineering from Institute of Software, Chinese Academy of Sciences, China in 2006. He was Post-doc fellow in School of Software, Tsinghua University, China from 2006-2007. He was a Post-doc fellow in the Software Engineering Laboratory at Pohang University of Science and Technology (POSTECH), South Korea, from 2007-2008. He was a lecturer of Biomedical Engineering Institute, Capital University of Medical Sciences, Beijing, China, from 2008 to 2009. He has been a member of IEEE since 2005. His research interests include: theoretical and empirical software engineering, model driven software development, knowledge engineering, ontology modeling, etc.

Christophe Cruz is an associate professor at the IUT of Dijon (Technological Institute of the University of Burgundy). He obtained a PhD in Computer Science at the University of Burgundy in 2004. He works on ontology modeling.

Christophe Nicolle is a full professor at the IUT of Dijon (Technological Institute of the University of Burgundy). He obtained a PhD in Computer Science at the University of Burgundy in 1996. His research area is the interoperability of information systems and more specially the problem of semantic heterogeneity.