# Knowledge Acquisition Based on Semantic Balance of Internal and External Knowledge

Vagan Y. Terziyan, Seppo Puuronen

Department of Computer Science and Information Systems,
University of Jyvaskyla, P.O.Box 35, FIN-40351, Jyvaskyla, Finland
e-mail: {vagan, sepi}@jytko.jyu.fi

**Abstract.** This paper presents a strategy to handle incomplete knowledge during acquisition process. The goal of this research is to develop formal tools that benefit the law of semantic balance. The assumption is used that a situation inside the object's boundary in some world should be in balance with a situation outside it. It means that continuous cognition of an object aspires to a complete knowledge about it and knowledge about internal structure of the object will be in balance with knowledge about relationships of the object with other objects in its environment. It is supposed that one way to discover incompleteness of knowledge about some object is to measure and compare knowledge about its internal and external structures in an environment. If there exist differences between the internal and the external semantics of an object, then these differences can be used to derive more knowledge about the object to make knowledge complete. The knowledge refinement process is done step-by-step as a continuous evolution of a knowledge base. Each step consists first automatic analysis of semantic balance which is then followed by attempts to derive knowledge that will balance differences between internal and external semantics of the object. This paper describes an algebra that is used to describe the internal and external semantics of an object and to derive unknown part of it. The results presented are mostly theoretical ones.

## 1 Introduction

This paper deals with a cognition strategy based on semantic model of world. It describes one refinement technique to handle incompleteness of knowledge in acquisition process. Knowledge base refinement is now one of the central problems of expert systems (Willkins [12]). It needs a fundamental research using basic concepts of philosophy and cognitive science. The main focus of this paper is to describe and apply one of the fundamental philosophic principles - "Balance in Nature" in terms of semantic networks to define the strategy of improving knowledge during a cognition process. The goal of this research is to develop formal metasemantic algebra that benefits the law of semantic balance, i.e. there should be balance between the internal and external semantics of an object in the possible world (*WORLD* in short onwards in this paper). If this semantic law holds in the *WORLD* and there exists any difference between the internal and the external semantics of an object, then this difference can be used to acquire more knowledge about the object. The refinement proceeds step-by-step as a continuous evolution of knowledge base, where each step includes two substeps: first substep makes automatic analysis of semantic balance and if the situation is not in balance then the second substep attempts to derive knowledge that will reestablish balance.

Knowledge base refinement as a method to improve an incorrect, inconsistent, and incomplete domain theory has also been suggested by Willkins [12]. His ODYSSEUS system refines knowledge bases of advanced rule-based systems. It learns by watching apprentice. His refinement program tries to construct an explanation of an observed action of an expert. Context of explanation allows to generate candidate of knowledge base repairs. ODYSSEUS system is designed for use with heuristic classification using hypothesis-directed reasoning. A processing stage prior to apprenticeship learning removes an inconsistent knowledge from the domain theory, which is responsible for deterioration of the performance of the system due to sociopathic interactions between elements of the domain theory. Sociopathicity implies that some kind of global refinement for the acquired knowledge is essential for machine learning.

Current books in formal semantics widely use approaches based on fundamental conceptual research in philosophy and cognitive psychology. For example Larson and Segal [6] give equal weight to philosophical, empirical, and formal discussions. They study a particular human cognitive competence governing the meanings of words and phrases. They argue that speakers have unconscious knowledge of the semantic rules of their language. Knowledge of meanings is both the semantics of domain attributes (properties and relations) and learning technology how to derive semantics of inconsistent and incomplete meanings.

During last several years one can see the growth of interest to semantic models of *World* (Li [7]). The reason seems to be in extremely fast development of global information networks. Study of large domains with numerous objects and groups of objects with relations requires possibilities to have closer considerations inside objects (their properties), outside objects (their external semantics), and both inside and outside considerations also for groups of objects. This kind of situations arises for example with WWW, the organization of which requires net-based semantic models and good technology of self-organization to handle problems of their complexity (Heylighen & Bollen [4]). One can interpret acquired knowledge only if "internal" part of it is in a conformity with "external" one. In other words these parts have to be in "balance". Phenomena of balance is very important in understanding problems related to knowledge (Schultz, Mareschal & Schmidt [10]). It was used by Kamimura [5] to minimise incompleteness of internal and external knowledge represented in neural networks. Balance has to be taken into account in cooperative modeling and machine learning, according to Morik [8] and DeJong [2], in systems control according to Sen & Jugo [11].

The main focus of this paper is to describe in formal way and apply the fundamental philosophic principle of balance between internal and external semantics of a domain object. We use and further develop the formalism of metasemantic algebra (Puuronen & Terziyan [9]), (Bondarenko, Grebenyuk & Terziyan [1]), (Grebenyuk et al. [3]) to describe internal and external semantics of any single or compound object in a network and the formal use of the law of semantic balance during the cognition process. Chapter 2 of this paper gives a short introduction to the metasemantic algebra. In chapter 3 the ways to formulate the internal and external semantics of an object is presented and chapter 4 introduces the law of semantic balance between the internal and external semantics of an object. Chapter 5 describes the stepwise process of knowledge refinement utilizing unbalanced situations and chapter 6 concludes with further research suggestions.

## 2 A Metasemantic Algebra

The metasemantic algebra was proposed in (Puuronen & Terziyan [9]) and further developed in (Bondarenko, Grebenyuk & Terziyan [1]). The basic elements of the algebra are objects and their relations with special operations upon semantic meanings of relations. In this chapter we will introduce the basic elements of semantic network and semantic operations.

### 2.1 A Semantic Network

Let $A_i$ be an atomic *object*. It can also have its internal structure but in relations it is considered as an atomic object. Let $L_k$ be a *semantic meaning* of *relation* between two objects or one object with itself. The second one corresponds to the *property* of the object. The *semantic predicate* $P$ is:

$$P(A_i, L_k, A_j) = \begin{cases} 1, \text{ if there is relation between} \\ \quad A_i \text{ and } A_j \text{ with meaning } L_k; \\ 0, \text{ otherwise.} \end{cases}$$

*Semantic network* $S$ is: $S = \bigwedge_{i,j,k} P(A_i, L_k, A_j)$, where $A_i$ is the source of the relation, $A_j$ is the object of the relation, and $L_k$ is the semantic meaning of the relation between those objects.

### 2.2 Semantic Constants and Operations

There are two *semantic constants*:

- semantic *ZERO* (notation - *IGN*): (it means a total ignorance about relationship between the source object and the target object):

$$\forall (A_i, A_j) \neg \exists L_k (P(A_i, L_k, A_j)) \Rightarrow P(A_i, IGN, A_j);$$

- semantic *UNIVERSE* (notation *SAME*): (it means a total knowledge about relationship between the source object and the target object).

There are two special relations *HAS_PART* and *PART_OF* which have their ordinary meanings. If it is true that:

$$P(A_i, HAS\_PART, A_j) \text{ or } P(A_j, PART\_OF, A_i),$$

then object $A_j$ is included in the object $A_i$. In the special case when an object is not part of any other object we call it as a (possible) *World*, i.e.

$$\forall A_i \neg \exists A_j (P(A_i, PART\_OF, A_j)) \Rightarrow A_i = World,$$

and in the special case when an object has no other object that is part of it, we call it as *Atom*, i.e.: $\forall A_i \neg \exists A_j (P(A_i, HAS\_PART, A_j)) \Rightarrow A_i = Atom$.

The ordinary *semantic operations* are:

- semantic inversion: $P(A_i, L_k, A_j) \equiv P(A_j, \tilde{L}_k, A_i)$;
- semantic addition:
$$P(A_i, L_k, A_j) \wedge P(A_i, L_n, A_j) \equiv P(A_i, L_k + L_n, A_j);$$
- semantic multiplication:
$P(A_i, L_k, A_m) \wedge P(A_m, L_n, A_j) \equiv P(A_i, L_k * L_n, A_j)$ where $i \neq j \neq m$.

In the graphical representation, objects are described by circles and relations with directed arcs leading from the source object to the target object. Object and relations that form an internal semantic structure of an object are presented inside the circle of that object. When the internal structure of an object is not under consideration it is not necessarily shown. The *WORLD* under consideration is presented by the outermost circle as in Figure 1 where the *WORLD* is $W$. It includes an object $A$ which internal structure is also presented in the figure. The object $A$ includes an object $A_2$ which internal structure is not presented. There is a relation $L_1$ between the objects $A_3$ and $A_1$ and the object $A_2$ has relationship with itself.

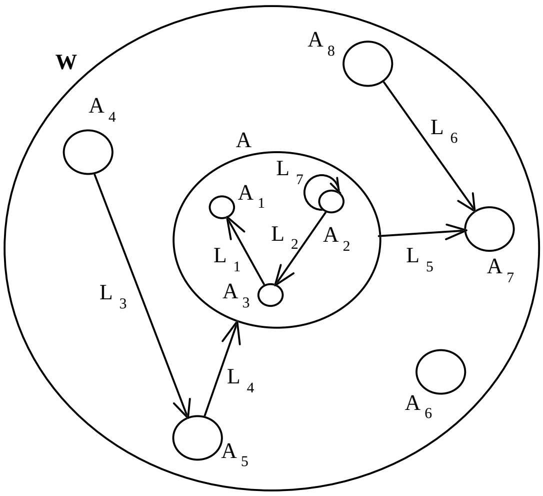


**Fig1.** The world *W* with its objects and relations.

## 3 The Internal and External Semantics

In this chapter we will derive the formulas of the semantic algebra which present the internal and external semantics of an object.

### 3.1 The Internal Semantics of an Object

We suppose that the internal semantics of an object can be defined using the components of the object and the relationships between those components. We further suppose that the quality of internal semantics behaves in monotonous way. It means that if any additional knowledge about the internal structure (components and their relations) of the object is achieved then the quality of the formulated internal semantics is never going worse. Thus by acquiring more and more knowledge about the object (its internal structure) we are able to achieve more and more complete internal semantics of this object.

We define that the internal semantics of an object $A_i$ is the semantic sum over all the possible paths between any pairs of objects $(A_j, A_k)$ included in the object $A_i$ plus the paths from each included object to itself. A path between any pair of objects $(A_j, A_k)$ includes successive relations (or their inverse) from the object $A_j$ to the object $A_k$ so that no relation (or its inverse) is taken twice. The only path where the same object is visited twice is the path from the object to itself. This later one guaranty that the properties of the included objects will be taken into account. Thus for a given object $A_i$ its internal semantics $E_{in}(A_i)$ is:

$$E_{in}(A_i) = \sum_{\substack{\forall j,k,j\leq k, \\ P(A_i,HAS\_PART,A_j) \\ P(A_i,HAS\_PART,A_k)}} L_{A_j - A_k}\,.$$

where $L_{A_j - A_k}$ is a path from $A_j$ to $A_k$.

In the case of Figure 1 the internal semantics of the object $A$ is:

$$E_{in}(A) = \tilde{L}_1 * \tilde{L}_2 + L_2 + \tilde{L}_1 + L_7\,.$$

The internal semantics corresponds the knowledge seen in Figure 2a.

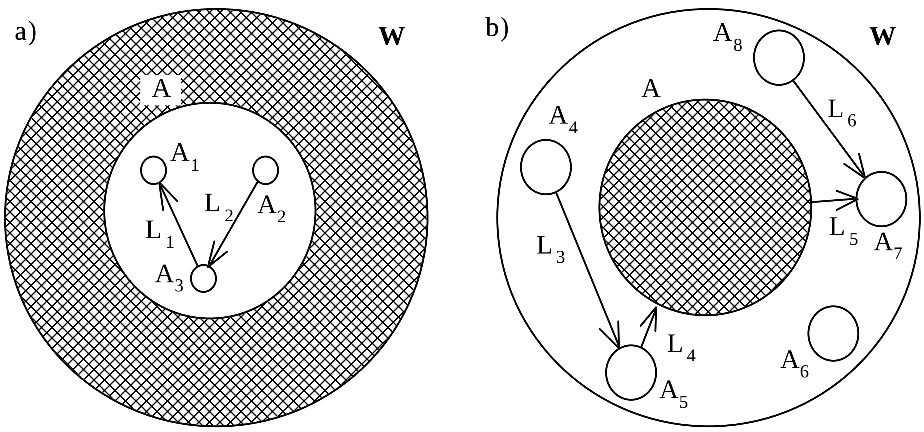


**Fig. 2.** a) The internal semantics of the object A in the possible world W, b) The external semantics of the object A in the possible world W.

### 3.2 The External Semantics of an Object

We suppose that the external semantics of an object *A* can be defined using the components of the possible world outside the object and the relationships between those components. We further suppose that the quality of external semantics behaves in monotonous way. It means that if any additional knowledge about the external structure (components and their relations) of the object is achieved then the quality of the formulated external semantics is never going worse. Thus by acquiring more and more knowledge about the world outside the object (its interaction with its environment) we are able to achieve more and more complete external semantics of this object.

We define that the external semantics of an object $A_i$ is the semantic sum over all the possible paths between any pairs of objects $(A_j, A_k)$ included in the world outside the object $A_i$ plus the paths from each included object to itself. The object $A_i$ belongs to the objects that participate in pairs but the path from the object $A_i$ to $A_i$ is not included. For a given object $A_i$ its external semantics is:

$$E_{ex}(A_i) = \sum_{\substack{\forall j,k,j \le k, j \ne k \ne i, \\ (A_j \in W/A_i) \vee (A_j = A_i), \\ (A_k \in W/A_i) \vee (A_k = A_i)}} L_{A_j - A_k}\,.$$

On the other hand $E_{ex}(A_i)$ is the internal semantics of the *World* when $A_i$ is taken as *Atom* (without noticing its internal structure). This gives a formula:

$$E_{ex}(A_i) = E_{in}(World / E_{in}(A_i)).$$

In the case of Figure 1 the external semantics of the object *A* is:

$$E_{ex}(A) = L_3 * L_4 + L_3 * L_4 * L_5 * \tilde{L}_6 + \\ + L_3 * L_4 * L_5 + L_4 + L_4 * L_5 * \tilde{L}_6 + \\ + L_4 * L_5 + L_5 * \tilde{L}_6 + L_5 + L_3 + \tilde{L}_6\,.$$

This external semantics corresponds the knowledge seen in Figure 2b.

## 4 The Law of Semantic Balance

Let us suppose that there exists a possible world where the ideal situation for an object $A_i$ is that its internal semantics (i.e. its internal structure = objects and their relations) and its external semantics (i.e. its properties when it interacts its environment) are in balance. In this ideal situation the law of semantic balance holds: $E_{in}(A_i) = E_{ex}(A_i)$. Usually, especially with knowledge bases, the ideal situation has not been achieved. Human knowledge about objects is almost always incomplete, and the knowledge base usually includes incompleteness, inconsistencies, and incorrectness. Sometimes we know more about the structure of an object than its external properties and sometimes on the contrary. Let $ign_{in}^{t}(A_i)$ be our ignorance about the internal semantics of the object $A_i$ at the time $t$ and let $ign_{ex}^{t}(A_i)$ be our ignorance about the external semantics of the object $A_i$ at the time t then according the law of semantic balance we can write

$$E_{in}^{(t)}(A_i) + ign_{in}^{(t)} = E_{ex}^{(t)}(A_i) + ign_{ex}^{(t)}.$$

## 5 Strategy of Knowledge Refinement

In this chapter we consider a strategy of improving incomplete or incorrect knowledge about some object during the acquisition process using the law of semantic balance. The strategy is shown in Figure 3.

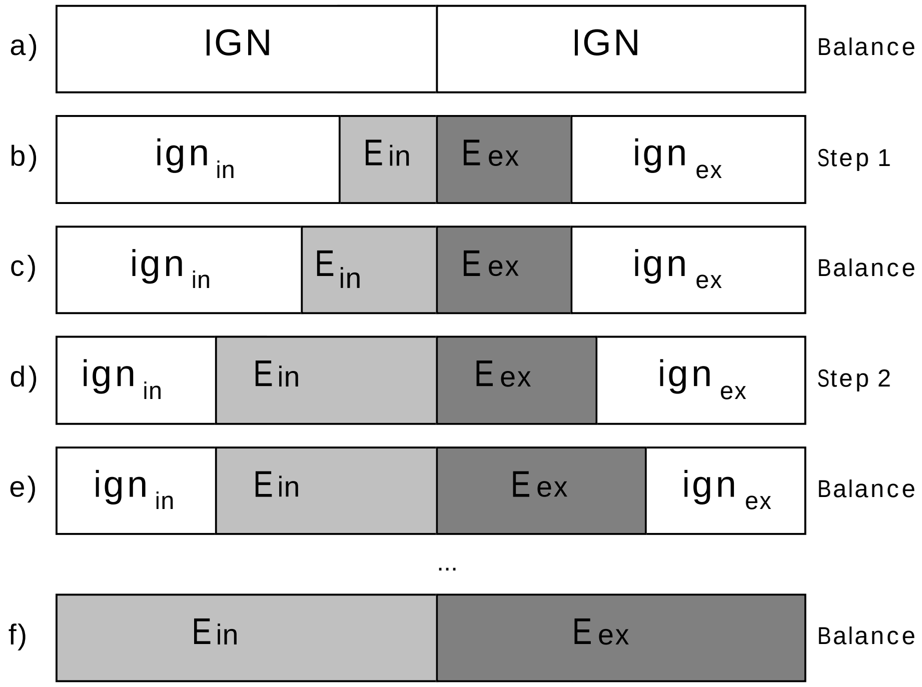


**Fig. 3.** The strategy of step-by-step knowledge refinement using the law of semantic balance.

Let us suppose that we have acquired some knowledge $E_{in}^{1}(A_i)$ about the internal semantics of the object $A_i$ and that we have acquired some knowledge $E_{ex}^{1}(A_i)$ about the external semantics of the object $A_i$. Let us assume that these two semantics are not in balance. Then we can try to make them in balance trying to remove some part of ignorance from either or both sides of the formula:

$$E_{in}^{(1)}(A_i) + ign_{in}^{(1)} = E_{ex}^{(1)}(A_i) + ign_{ex}^{(1)}.$$

If this equation succeeds, at least partially, then we receive another amount of knowledge $E_{in}^2(A_i)$ about the internal semantics of the object $A_i$ and another amount of knowledge $E_{ex}^2(A_i)$ about the external semantics of the object $A_i$. If these two semantics are not in balance or if some outer knowledge source gives extra knowledge that makes them unbalance again, then we try to make them in balance trying to remove some part of ignorance using the same formula as above and so on (Figure 3).

## 6 Conclusion

We have discussed some aspects of using semantic balance in knowledge acquisition. It uses as a basic measure of unbalance the difference between the internal and external semantics of an object. These semantics are expressed using formal metasemantic algebra. The solution of equations of this algebra needs further development. Nevertheless the principle of semantic balance seems to be a good way to understand the dynamics of knowledge.

## References


1. Bondarenko, M.; Grebenyuk, V.; and Terziyan, V. 1993. Reasoning Based on the Algebra of Semantic Relation. *Pattern Recognition and Image Analysis* 3(4):488-499.
2. DeJong, G.; and Bennett, S. 1993. Permissive Planning: A Machine Learning Approach to Linking Internal and External Worlds, *AAAI Press/MIT Press*, 508-513. Menlo Park.
3. Grebenyuk, V.; Kaikova, H.; Terziyan, V.; and Puuronen, S. 1996. The Law of Semantic Balance and its Use in Modeling Possible Worlds. In *STeP-96 - Genes, Nets and Symbols*, 97-103. Publications of the Finnish AI Society.
4. Heylighen, F.; and Bollen, J. 1996. The World Wide Web as a Super-Brain: From Metaphor to Model. In Proceedings of the 13-th European Meeting on Cybernetics and Systems Research, 917-922. Austrian Society for Cybernetic Studies.
5. Kamimura, R. 1992. Generation of Inhibitory Connections to Minimize Internal and External Entropy. In Alexander I. and Taylor J. (eds.), Artificial Neural Networks, 2, 969-972. North Holland, Amsterdam.
6. Larson, R.; and Segal, G. 1995. *Knowledge of Meaning. An Introduction to Semantic Theory*, Mass.: A Bradford Book.
7. Li, L. 1994. Possible World Semantics and Autoepistemic Reasoning. *Artificial Intelligence* 71(2).
8. Morik, K. 1993. Balanced Cooperative Modeling. *Machine Learning* 11(2/3).
9. Puuronen, S.; and Terziyan, V. 1992. A Metasemantic Network. In *STeP-92 - New Directions in Artificial Intelligence*, 136-143. Publications of the Finnish AI Society.
10. Schultz, T.R.; Mareschal, D.; and Schmidt, W.C. 1994. Modeling Cognitive Development on Balance Scale Phenomena. *Machine Learning* 16(1/2).
11. Sen, M. de la; and Jugo, J. 1994. Robust Control for a Class of Systems with Internal and External Point Delays, *Cybernetics and Systems '94,* 191-198. World Scientific Publishing.
12. Willkins, D.C. 1993. Knowledge Base Refinement as Improving an Incorrect and Incomplete Domain Theory. In Buchanan B.G., Wilkins D.C. (eds.), *Readings in Knowledge Acquisition and Learning*, 631-641. Morgan Kaufmann, Calif.