Machines with good sense: How can computers become capable of sensible reasoning?

Junia C. Anacleto, Ap. Fabiano Pinatti de Carvalho, Eliane N. Pereira, Alexandre M. Ferreira, Alessandro J. F. Carlos

Abstract Good sense can be defined as the quality which someone has to make sensible decisions about what to do in specific situations. It can also be defined as good judgment. However, in order to have good sense, people have to use common sense knowledge. This is not different to computers. Nowadays, computers are still not able to make sensible decisions and one of the reasons is the fact that they lack common sense. This paper focuses on OMCS-Br, a collaborative project that makes use of web technologies in order to get common sense knowledge from a general public and so use it in computer applications. Here it is presented how people can contribute to give computers the knowledge they need to be able to perform common sense reasoning and, therefore, to make good sense decisions. In this manner, it is hoped that software with more usability can be developed.

1 Introduction

Until nowadays computer are not capable of understanding about ordinary tasks that people perform in their daily life. They cannot reason about simple things using good sense as a person can do, and, therefore, they cannot help their users as they could if they had the capacity of making good judgment about the users' needs.

Since late 1950s, Artificial Intelligence (AI) researches have been looking for ways to make computers intelligent so that they could help their users in a better way. Part of those researchers believes that, in order to be intelligent, computers should first get the knowledge about human experiences, which involves knowledge about spatial, physical, social, temporal, and psychological aspects of typical everyday life. The set

---

Junia C. Anacleto, Ap. Fabiano Pinatti de Carvalho, Eliane N. Pereira, Alexandro M. Ferreira, Alessandro J. F. Carlos

Advanced Interaction Laboratory (LIA)
Computing Department - Federal University of São Carlos (DC/UFSCar)
Rod Washington Luis, Km 235 – São Carlos – SP – 13565-905
{junia, fabiano, eliane_pereira, alexandre_ferreira, alessandro_carlos}@dc.ufscar.br
+55 16 3351-8618

of these kinds of knowledge, which is shared by most of people who have the same cultural background, is called common sense [1], [8], [11].

Actually common sense knowledge is very important to reach good sense because people use it to make their judgments. For example, in order to judge as wrong the attitude of a child badly responding to his parents, people have to consider some common sense facts such as “a child must respect older people”, “parents are older than their children”, “responding in a bad way is not respectful”, and so on.

Since common sense is essential to reach good sense, how can this knowledge be provided to computers? One idea is to construct machines that could learn as a child, observing the real world. However, this approach was discarded after Minsky’s and Papert’s experience [15] of building an autonomous hand-eye robot, which should perform simple tasks like building copies of children’s building-block structures. With this experience, they realized that it would be necessary innumerous short programs to give machines human abilities as cognition, perception and locomotion.

Another idea is to build a huge common sense knowledge base, to store it in computers and to develop procedures which can work on that knowledge. This seems an easier approach; nevertheless there are big challenges that must be won in order to get it [2], [7]. The first challenge of the second idea is to build the common sense knowledge base, since it is believed that to cover the human common sense it is necessary billions of pieces of knowledge such as knowledge about the world, myths, beliefs, etc. [1], [9], [10], and it is know that common sense is cultural dependent [1], [2]. Other challenges are presented in further sections.

Talking about building a large scale common sense knowledge base and developing applications capable of common sense reasoning, one attempt of this is the CYC Project, idealized by Douglas Lenat and that has been under development since 1984 [9]. In this project, Knowledge Engineers work on data that is gotten by interviewing people, and populate the project common sense knowledge base, storing the knowledge in a specific language, CycL. This approach has been proved to be very expensive, since nowadays the sum expended with the project exceeds tens of millions dollars [16]. In this way, CYC has been working on other alternatives [16].

Another attempt to build the desired common sense knowledge base and to use it in computer applications is the OMCS (Open Mind Common Sense) project [17], which takes into account the fact that every ordinary people has the common sense that computers lack and, so, everyone can help to construct the base. In this project the web technologies play a very important role in the knowledge base building. In order to get common sense facts it was developed web sites where anyone who knows how to write in a specific language – there are many versions of the project, each one in a language, such as English, Portuguese, Mexican and Korean – can subscribes himself and contribute by entering statements in natural language that originates a semantic network which is used by computer application.

This paper focuses on the OMCS-Br (Brazilian Open Mind Common Sense) project and its approaches to give common sense to computers. It is organized as follows: section 2 goes over some challenges that were faced since OMCS-Br has been under development at the Advanced Interaction Laboratory (LIA) from Federal

University of São Carlos, Brazil; section 3 brings some accomplishments of the project; and section 4 presents some conclusions and points to future works.

2 Challenges of getting and using common sense

Providing computers with common sense knowledge is an old dream of some AI researchers. In 1959, McCarthy was already concerned about the importance of giving this kind of knowledge to machines in order to make them intelligent [14]. Actually, there are those, as Marvin Minsky, who believes that the true intelligence with which computers should be supplied, lays on this kind of knowledge [15]. In spite of that, few projects have been developed to the purpose of reaching this dream. This is because there are difficult issues to deal with, as the ones experienced by OMCS-Br and explored in this paper.

First of all, to build a robust inference system based on common sense knowledge, it is necessary to construct a huge knowledge base [9], [10], [7]. However, what can be considered a huge knowledge base? Concerning those projects that collect common sense using natural language sentences, how many sentences should be necessary to cover the whole human knowledge? Furthermore, since it is known that common sense changes as time goes by, how long does it take to build the desired knowledge base? These are some questions that still have no answers, and maybe it was what has leaded some AI researchers not to invest on building common sense knowledge bases.

Nevertheless, OMCS-Br researchers believe that, whereas there is still no huge common sense knowledge base, it is possible to make machines more helpful and quite intelligent with a bit of common sense knowledge gotten from web contributors, as [10] showed to be possible. However, other questions rise up, as the ones related to the knowledge base quality. How can collaborators be guided in order to enter sentences related to the several kinds of knowledge which compose people's common sense? How should redundancy be treated? What about orthographic mistakes? A last question concerning the knowledge base construction: how can users be motivated to contribute on the website?

Now talking about the knowledge pre-processing which is necessary in order to use the acquired knowledge in computer application [9] [11], natural language has several syntactic structures. How should sentences be managed in order to get a better use of the knowledge they express? Which natural language parser should be used?

These are some questions which OMCS-Br has faced with since it has been launched. The approaches adopted by the project to some of them are presented in the next section.

3 The OMCS-Br Project accomplishments

Notwithstanding challenges, OMCS projects have been working to overcome all of them. Here it is presented how OMCS-Br has been approaching some issues previously mentioned. To begin with the knowledge base building, OMCS-Br adopts template-based activities which guide users in such a way that they can contribute with different kinds of knowledge. The templates are semi-structured sentences in natural language with some lacunas that should be filled out with the contributors' knowledge so that the final statement corresponds to a common sense fact. They were planned to cover those kinds of knowledge previously mentioned and to get pieces of information that will be used further to give applications the capacity of common sense reasoning. The template-based approach makes easier to manage the knowledge acquired, since the static parts are intentionally proposed to collect sentences which can be mapped into first order predicates, which composes the project's semantic network. In this way, it is possible to generate extraction rules to identify the concepts present in a statement and to establish the appropriate relation-type between them. In OMCS projects, there are twenty relation-types, used to represent the different kinds of common sense knowledge, as it is presented in [11].

Those templates have a static and a dynamic part. The dynamic part is filled out by a feedback process that uses part of sentences stored in the knowledge base of the project to compose the new template to be presented. Figure 2 exemplifies how the feedback process works. At the first moment the template "You usually find a ___________ in a **chair**" of the activity *Location* is presented to a contributor – the templates bold part is the one filled out by the feedback system. In the example, the contributor fills out the sentence with the word "*screw*". Then, the sentence *"You usually find a screw in a chair"* is stored in the OMCS knowledge base. At the second moment, the template "A **screw** is used for __________" of the activity *Uses* is shown to another contributor. Note that the word *screw* entered at the first moment is used to compose the template presented at the second moment.

Figure 1. Example of the OMCS-Br feedback process

The feedback process used in OMCS-Br website was planned in order to allow varied templates to be generated so that users are able to contribute on several subjects and do not get bored with always filling out the same sentence.

Still related to the feedback process, considering that the sentences stored in the knowledge base will be used to compose templates that will be shown to other contributors, it is important to provide a way through what it could be selected the sentences that should be used by the feedback process. Thinking in this need, it was developed in OMCS-Br an on-line review system, which can be accessed just by the ones who have administrator privileges, where the sentences are selected to be or not to be used by the feedback process.

In order to perform the review, it was defined some rules to assure that common sense knowledge would not be discarded. The rules adopted in the review process are the following:

1. Sentences generated from a template that was filled out with a set of character without any meaning in Brazilian Portuguese are rejected. For example – if someone fills out a template with "*dafasdfasd*" the sentence is rejected;
2. Sentences with errant spelling, e.g., sentences that were filled out with words that are written orthographically wrong, are rejected;
3. Sentences generated by a template which was filled out differently from the default defined by the Knowledge Engineers to that activity, are accepted, but the entry is not used in feedback process. This happened for example, when the Knowledge Engineer defined that the default entry to a template is a noun phrase but the contributor filled it out with a verbal phrase. The entry is accepted, if all words are orthographically correct. The justification to this approach is that if the entry is accepted to the feedback process, it will be generated templates syntactically incorrect.
4. Sentences generated by a template that was filled out with bad words are accepted, but this entry is not used by the feedback process.

It is worth pointing out that during the review process the reviewer is not allowed to judge the semantic of a sentence. That is because it does not matter if a sentence seems strange in meaning or if it has already been scientifically proved as wrong. Common sense knowledge does not match scientific knowledge necessarily. Since a sentence is accepted as true by the most people who share the same cultural background, it is considered as a common sense sentence. Because of that reviewers are not allowed to judge if a sentence is common sense sentence or not.

Besides the templates about general themes such as those about *"things"* which people deal with in their daily life, *"locations"* where things are usually found and the common *"uses"* of things, there are also, in the Brazilian project website, templates about three specific domains: **health**, **colors** and **sexual education**. They are domains of interest to the researches that are under development in the research group which keeps the project [5] [4] [2]. This approach is only used in Brazil and it was adopted taking into account the necessity of making faster the collection of common sense knowledge related to those domains. The specific-domain templates were defined with the help of professionals of each domain. They were composed with some specifics words which instantiate the templates of general themes, in order to guide users to contribute with sentences related to a domain. Table 1 shows the accomplishments that OMCS-Br has gotten with that approach.

Table 1. Contributions on specific domains in OMCS-Br

Domain	Number of contributions	Period of collection
Health	6505	about 29 months
Colors	8230	about 26 months
Sexual Education	3357	about 21 months

The numbers of contributions in each domain can seem to be irrelevant, however, considering the only 2 facts about AIDS found in the knowledge base before creating the theme Sexual Education, it can be noticed the importance of domain-contextualized templates in order to make fast the collection of statements related to desired domains.

Another accomplishment of the OMCS-Br is related to the variety of contributor profiles. Nowadays there are 1499 contributors registered in the project site of which 19.33% are women and 80.67% are men. The most part of contributors (72.80%) is from Brazil South-east area, followed by the South area (15.25%). Those numbers point to the tendency that is proved by geographic sciences, which present the South-east and South area as being the most developed areas of Brazil. Considering that, it is perfectly understandable that, being well developed areas, their inhabitants have easier access to the Internet. Table 2 and Table 3 present other characteristics of OMCS-Br contributors.

Table 2. Percentage of contributors by age group

Age group	Percentage
Younger than 12 years	0.75 %
13 – 17	20.51 %
18 – 29	67.36 %
30 – 45	9.88 %
46 – 65	1.22 %
Older than 65 years	0.28 %

Table 3. Percentage of contributors by school degree

School degree	Percentage
Elementary school	2.21 %
High school	18.17 %
College	65.86 %
Post-Graduation	4.52 %
Master Degree	7.04 %
Doctorate Degree	2.21 %

Another conquest of OMCS-Br is the amount of contributions. Within two years and a half of project, it was gotten more than 174.000 sentences written in natural language. This was possible thanks the web technology and the marketing approach adopted by LIA. As the project was released in Brazil in 2005, it was realized that the knowledge base would rise up significantly just when there were an event that put the project in evidence. Figure 2 demonstrates this tendency.

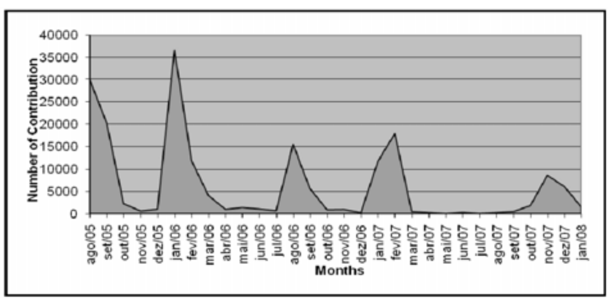

Figure 2 OMCS-Br knowledge base tendency of growing up

It can be noticed in Figure 2 that the periods where the knowledge base grew up significantly were from August to October 2005, from January to March 2006, from August to October 2006, from January to February 2007 and from November to December 2007. This is an interesting fact, because those jumps in the knowledge base just followed some marketing appeals performed by LIA. In the first one, LIA got published some articles in some newspapers of national coverage telling people about the project and asking for people contribution. After had those articles printed, the OMCS-Br knowledge base reach the number of 50.000 sentences. Three months later, the knowledge base established and passed to grow up very slowly.

Thinking of having another jump in the knowledge base size, it was released in the later January 2006 a challenge associated to the Brazilian carnival. In that challenge, it was offered little gifts as prizes to the three first collaborators that contributed with the most number of sentences in the site activities. The winners received T-Shirts of the OMCS-Br Project and pens of MIT. The challenge was announced among the project contributors, which received an e-mail telling about it. The announcement was also posted in the Ueba website (www.ueba.com.br), a site of curiosities which target public is people interested in novelties. As it can be noticed, the knowledge base size had a jump as soon as the challenge was launched. The same approach was used in August 2006, January 2007 and December 2007.

Although the approach has gotten a good response from the contributors in the first three challenges, it can be noticed in Figure 2 that this approach is becoming inefficient. Thinking about keeping the knowledge base growing up, it is under development some games, following project contributors' suggestions, in order to make the collection process funnier and more pleasant.

Besides the knowledge base growth, another important issue in OMCS-Br is the pre-processing of the sentences stored in the knowledge base. As the knowledge is collected in natural language, it might be put in a computational notation in order to be used in computer application. The knowledge representation adopted in OMCS-Br is a semantic network.

After being generated in the *extraction* process, i.e. the process which extracts the semantic network nodes from the natural language statements stored in the knowledge base and relates them through first order predicates, the network nodes are submitted to a *normalization* process. Since the sentences collected in the site can vary in their morphology, it is needed to manipulate those sentences in order to increase the semantic network connectivity.

In order not to have inflected concepts, which means same words varying in number, tense, etc, separated in the semantic network, a set of heuristics is applied on the contributions so that they are grouped in a single node of the semantic network.

The normalization process in OMCS-Br is performed using Curupira [13], a syntactic parser for Brazilian Portuguese. However, as the parser does not strip the sentence inflectional morphology, it was developed a Python module to normalize the nodes. For this purpose, it is used the inflectional dictionary developed in the UNITEX-PB Project [18], which has all inflectional forms of Brazilian Portuguese morphological classes.

The module works in 3 steps. First of all, each sentence token is tagged using the Curupira parser. Afterward, articles and cardinal numbers are taken off – proper nouns are kept in original form. Special Portuguese language structures are treats. For instance, the *ênclise* structure, which is a case of pronominal position where the pronoun is concatenated after the verb, is stripped from the sentences and the verb is put in the infinitive form. For example, the verb "*observá-la*" ("observe it") is normalized to "*observar*" ("to observe"). Overall, each tagged token is normalized by searching its normal form in the inflectional dictionary used. In this way, sentences that were separated by morphological variations, like "*comeria maçãs*" ("would eat apples") and "*comendo uma maçã*" ("eating an apple"), are reconciled during the normalization process generating the normalized expression "*comer maçã*" ("to eat apple").

In order to check the connectivity of the network generated using and not using the normalization process a test was performed. The results of this measurement are presented in Table 4.

Table 4 – Effects of the normalization process on the OMCS-Br semantic network structure

	non-normalized	normalized	normalized/ non-normalized
nodes	36,219	31,423	- 13.24 %
relations	61,455	57,801	- 5.95 %
average nodal edge-density	4.4643	3.3929	+ 31.57 %

These results can be interpreted as follows: regarding the number of nodes and relations, they were decreased after the normalization process. This confirms the tendency that the normalization process makes reconciliations between morphological variations, and thus unifies them.

Another result that can be inferred examining the connectivity of semantic network is that the nodal edge-density has increased more than 30%, which demonstrates that the normalization process improves the connectivity of nodes.

Other strategy to improve the connectivity of the network is to extract new relations from the original relations. This is made applying a set of heuristic inferences over the original relations nodes. The relations generated by these heuristics are K-Line relations, a kind of relation based on Minsk's Theory about the contextual mechanism in memory [15].

One of these heuristics identifies whether a node is composed by more than a word, finds the node components variations based on grammar patterns and establishes "ThematicKLine" relations between the variations which do not have any word in common. For example, in the node "*pote de mel na mesa*" ("honey jar on the table") it is found the following variations: "*pote de mel*" ("honey jar"), "*pote*" ("jar"), "*mel*" ("honey") e "*pote na mesa*" ("jar on the table"). So, it is generated the following ThematicKLine:

```
(ThematicKLine 'pote de mel' 'mesa')
(ThematicKLine 'pote na mesa' 'mel')
(ThematicKLine 'pote' 'mesa')
(ThematicKLine 'pote' 'mel')
(ThematicKLine 'mesa' 'mel')
```

Another heuristic considers the nominal adjuncts in a node. In Portuguese, the nominal adjunct is a phrase accessory term that delimits or specifies a noun, and can be composed by a noun followed by an adjective. With this construction, it is created "SuperThematicKLine" relations, which establish generalization/specialization relation between the nodes. This relation links the entire structure to the stripped adjective structure. For example, from the expressions "*sala grande*" ("big room") it is created the following relations:

```
(SuperThematicKLine 'sala grande' 'sala')
(SuperThematicKLine 'sala grande' 'grande')
```

In this way, related terms are linked one another in the semantic network which became consequently more connected.

These are the approaches which are used by the OMCS-Br project. The next section presents some conclusions on providing common sense to computers so that they can make sensible reasoning and points to some projects which are under development using the architecture of this project.

4 Conclusions and Future works

This paper presented the approaches adopted by OMCS-Br to collect common sense knowledge from a general public and use it in computer applications. The project has been working on three fronts to make possible the development of applications which are capable of common sense reasoning. It is believed that, giving computer this ability is a step on getting machines which can act with good sense. In this way, it would be possible to construct applications which can support their users in a better way, offering a contextualized help, according to the common sense knowledge which the machines were provided with.

A research developed at LIA has pointed to the fact that OMCS-Br knowledge bases store cultural differences as it is presented in [1] and in [2]. As future work, it is intended to invest in the development of applications with intelligent interfaces. Those interfaces would take into account the cultural context, since it is known that cultural differences impacts directly in the user interface [12]. Considering common sense knowledge, applications could offer an interaction instantiated to that cultural background. Another research developed at the laboratory is related to using common sense knowledge to support teachers to plan learning activities [6]. It is being investigated, how common sense knowledge can be used by teachers in order to make them concerned about the previous knowledge of its students, about the misconceptions that should be approached during a learning activity, since common sense register myths, believes and procedures of the daily life, and so on. Also common sense reasoning has been integrated to *Cognitor*, an authoring tool developed at LIA, whose main purpose is to support the development of learning material to be delivery electronically [3]. There is another research related to how common sense reasoning can be used in the development of educational games [4], which allow teachers to use common sense knowledge in order to contextualize the learning process.

Actually, there are lots of challenges to be won in order to reach the dream of make machines capable of common sense reasoning and, consequently, good sense reasoning. The OMCS-Br group is concerned about the innumerous challenges which they might deal with and it has been looking for solutions that can lead to the success of the project as a whole.

Acknowledgments

We thank FAPESP and CAPES for partially support this project. We also thank NILC that made Curupira *dll* available to us.

References

1. Anacleto, J. C.; Lieberman, H.; Tsutsumi, M.; Neris, V. P. A.; Carvalho, A. F. P.; Espinosa, J.; Zem-Mascarenhas, S.; Godoi, M. S. Can common sense uncover cultural differences in computer applications?. In *Artificial Intelligence in Theory and Practice – WCC 2006*, vol. 217, Bramer, M. (Ed). Berlin: Springer-Verlag, 2006. p. 1-10.
2. Anacleto, J. C.; Lieberman, H.; Carvalho, A. F. P.; Neris, V. P. A.; Godoi, M. S.; Tsutsumi, M.; Espinosa, J.; Talarico Neto, A.; Zem-Mascarenhas, S. Using common sense to recognize cultural differences. In *Advances in Artificial Intelligence – IBERAMIA-SBIA 2006*, LNAI 4140, Sichman, J. S. et al. (Eds). Berlin: Springer-Verlag, 2006. p. 370-379.
3. Anacleto, J. C.; Carlos, A. J. F.; Carvalho, A. F. P. de; Godoi, M. S. Using Common Sense Knowledge to Support Learning Objects Edition and Discovery for Reuse. In *Proc of the 13th Brazilian Symposium on Multimedia and the Web (Webmedia 2007)*, 2007, Gramado – Brazil. New York: ACM Press, 2007. p. 290-297.
4. Anacleto, J. C.; Ferreira, A. M.; Pereira, E. N.; Carvalho, A. F. P. de; Fabro, J. A. Culture Sensitive Educational Games Considering Commonsense Knowledge. In *Proc. of 10th International Conference on Enterprise Information System* (ICEIS 2008), 2008, Spain. In press.
5. Carvalho, A. F. P. de; Anacleto, J. C.; Zem-Mascarenhas, S. Learning Activities on Health Care Supported by Common Sense Knowledge. In: *Proc of the 23rd Symposium on Applied Computing*, 2008, Fortaleza – Brazil. New York: ACM Press, 2008 p. 1385-1389.
6. Carvalho, A. F. P. de; Anacleto, J. C.; Zem-Mascarenhas, S. Planning Learning Activities Pedagogically Suitable by Using Common Sense Knowledge. In *Proc of the 16th International Conference on Computing (CIC 2007)*, 2007, Mexico City. p. 1-6.
7. Denison , D.C. Guess who's smarter. *Boston Globe Online*, 25 mar. 2003.
8. Lenat, D. B. CYC: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, New York, v. 38, n. 11, Nov. 1995.
9. Lenat, D. B.; Guha, R. V.; Pittman, K.; Pratt, D.; Shepherd, M. Cyc: toward programs with common sense. *Communications of the ACM*, New York, v. 33, n. 8, aug. 1990.
10. Lieberman, H.; Liu, H. Singh, P.; Barry, B. Beating common sense into interactive applications. *AI Magazine*, v. 25, n. 4, p. 63-76, 2004.
11. Liu, H.; Singh P. ConceptNet: a practical common sense reasoning toolkit. *BT Technology Journal*, v. 22, n. 4, p. 211-226, 2004.
12. Marcus, A. Culture class vs. culture clash. *Interactions*, New York, v. 9, n. 3, p. 25-28, mar. 2002.
13. Martins, R. T.; Hasegawa, R.; Nunes, M.G.V. Curupira: a functional parser for Brazilian Portuguese. In *Proc of the 6th Computational Processing Of The Portuguese Language, International Workshop (PROPOR 2003)*, LCNS 2721, Mamede, N. J et al. (Eds). Berlin: Springer-Verlag, 2003.
14. McCarthy, J. Programs with Common Sense. In *Proc of the Teddington Conference On Mechanization Of Tought Processes*, 1959.
15. Minsky, M. *The Society of Mind*. New York: Simon and Schuster, 1986.
16. Panton, K.; Matuszek, C.; Lenat, D. B.; Schneider, D.; Witbrock, M. Siegel, N. Shepard, B. Common Sense Reasoning – From Cyc to Intelligent Assistant. In *Ambient Intelligence in Everyday Life*. New York: Spring-Verlag, LNAI 3864, 2006. pp. 1-31.
17. Singh, P. The OpenMind Common sense project. *KurzweilAI.net,* 2002.
18. UNITEX-PB– http://www.nilc.icmc.usp.br:8180/ unitex-pb/index.html