Open Mind Common Sense

Open Mind Common Sense (**OMCS**) is an artificial intelligence project based at the Massachusetts Institute of Technology (MIT) Media Lab whose goal is to build and utilize a large commonsense knowledge base from the contributions of many thousands of people across the Web.

Since its founding in 1999, it has accumulated more than a million English facts from over 15,000 contributors in addition to knowledge bases in other languages. Much of OMCS's software is built on three interconnected representations: the natural language corpus that people interact with directly, a semantic network built from this corpus called **ConceptNet**, and a matrix-based representation of ConceptNet called **AnalogySpace** that can infer new knowledge using dimensionality reduction.[1] The knowledge collected by Open Mind Common Sense has enabled research projects at MIT and elsewhere.

History

The project was the brainchild of Marvin Minsky, Push Singh, Catherine Havasi, and others. Development work began in September 1999, and the project was opened to the Internet a year later. Havasi described it in her dissertation as "an attempt to ... harness some of the distributed human computing power of the Internet, an idea which was then only in its early stages."[2] The original OMCS was influenced by the website Everything2 and its predecessor, and presented a minimalist interface that was inspired by Google.

Push Singh was slated to become a professor at the MIT Media Lab to lead the Common Sense Computing group in 2007 until his suicide on Tuesday, February 28, 2006.[3]

The project is currently run by the Digital Intuition Group at the MIT Media Lab under Catherine Havasi.

Database and website

There are many different types of knowledge in OMCS. Some statements convey relationships between objects or events, expressed as simple phrases of natural language: some examples include "A coat is used for keeping warm", "The sun is very hot", and "The last thing you do when you cook dinner is wash your dishes". The database also contains information on the emotional content of situations, in such statements as "Spending time with friends causes happiness" and "Getting into a car wreck makes one angry". OMCS contains information on people's desires and goals, both large and small, such as "People want to be respected" and "People want good coffee".[1]

Originally, these statements could be entered into the Web site as unconstrained sentences of text, which had to be parsed later. The current version of the Web site collects knowledge only using more structured fill-in-the-blank templates. OMCS also makes use of data collected by the Game With a Purpose "Verbosity".[4]

In its native form, the OMCS database is simply a collection of these short sentences that convey some common knowledge. In order to use this knowledge computationally, it has to be transformed into a more structured representation.

ConceptNet

ConceptNet is a semantic network based on the information in the OMCS database. ConceptNet is expressed as a directed graph whose nodes are concepts, and whose edges are assertions of common sense about these concepts. Concepts represent sets of closely related natural language phrases, which could be noun phrases, verb phrases, adjective phrases, or clauses.[5]

ConceptNet is created from the natural-language assertions in OMCS by matching them against patterns using a shallow parser. Assertions are expressed as relations between two concepts, selected from a limited set of possible relations. The various relations represent common sentence patterns found in the

OMCS corpus, and in particular, every "fill-in-the-blanks" template used on the knowledge-collection Web site is associated with a particular relation.[5]

The data structures that make up ConceptNet were significantly reorganized in 2007, and published as ConceptNet 3.[5] The Software Agents group currently distributes a database and API for the new version 4.0.[6]

In 2010, OMCS co-founder and director Catherine Havasi, with Robyn Speer, Dennis Clark and Jason Alonso, created Luminoso, a text analytics software company that builds on ConceptNet.[7][8][9][10] It uses ConceptNet as its primary lexical resource in order to help businesses make sense of and derive insight from vast amounts of qualitative data, including surveys, product reviews and social media.[7][11][12]

Machine learning tools

The information in ConceptNet can be used as a basis for machine learning algorithms. One representation, called AnalogySpace, uses singular value decomposition to generalize and represent patterns in the knowledge in ConceptNet, in a way that can be used in AI applications. Its creators distribute a Python machine learning toolkit called Divisi [13] for performing machine learning based on text corpora, structured knowledge bases such as ConceptNet, and combinations of the two.

Comparison to other projects

Other similar projects include Never-Ending Language Learning, Mindpixel (discontinued), Cyc, Learner, SenticNet, Freebase, YAGO, DBpedia, and Open Mind 1001 Questions, which have explored alternative approaches to collecting knowledge and providing incentive for participation.

The Open Mind Common Sense project differs from Cyc because it has focused on representing the common sense knowledge it collected as English sentences,

rather than using a formal logical structure. ConceptNet is described by one of its creators, Hugo Liu, as being structured more like WordNet than Cyc, due to its "emphasis on informal conceptual-connectedness over formal linguistic-rigor".[14]

There is also the Brazilian initiative, named Open Mind Common Sense in Brazil (OMCS-Br), led by the Advanced Interaction Lab at Federal University of São Carlos (LIA-UFSCar). This project started in 2005, in collaboration with the Software Agents Group at the MIT Media Lab, the main goal is to collect common sense stated in Brazilian Portuguese and use it to develop culturally sensitive software applications based on extracting cultural profiles' knowledge from ConceptNet. Hopefully this will help developers and users with a culturally contextualized content software, making the final applications more flexible, adaptive, accessible and usable. The main applications' focuses are education and healthcare.

See also

- Attempto controlled English

- Never-Ending Language Learning

- Mindpixel

- ThoughtTreasure

- Semantic Web

- DBpedia

- Freebase (database)

- yago (database)

References

1. Robyn Speer, Catherine Havasi, and Henry Lieberman. AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge   Archived 2010-07-09 at the Wayback Machine. AAAI 2008.

2. Catherine Havasi. Discovering Semantic Relations Using Singular Value Decomposition Based Techniques. Ph.D Thesis, Brandeis University June 2009.

3. MIT News Office (2006-03-08). "Memorial service slated tomorrow for Pushpinder Singh" ⁄ ̔̕G ̕G̔̍̕D̔G ̍̕. Retrieved 2009-10-07.

4. "Profile for verbosity" . Open Mind Commons Sense. Archived from the original on 2010-06-25.

5. Catherine Havasi, Robyn Speer and Jason Alonso. ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge . Proceedings of Recent Advances in Natural Language Processing, 2007. try ConceptNet 3:... Archived 2015-05-29 at the Wayback Machine

6. Commonsense Computing Initiative (2009-02-24). "ConceptNet API in Launchpad" . Retrieved 2009-10-07.

7. Lohr, Steve (27 June 2014). "The U.S.-Germany Match Through a Social Media Lens" . fi̔Gy̆ ̆u̇r ̆v̍̕ ̔GẹN ̔GY̆. Retrieved 3 March 2015.

8. Rusli, Evelyn (14 April 2014). "Firms Use Artificial Intelligence to Tap Shoppers' Views" . The Wall Street Journal. Retrieved 3 March 2015.

9. Alba, Davey (12 February 2015). "The Startup That Helps You Analyze Twitter Chatter in Real Time" . Wired. Retrieved 3 March 2015.

10. Noyes, Katherine (11 February 2015). "Luminoso to enterprises: Here's what all that chatter really means" . PC World. Retrieved 3 March 2015.

11. Miller, Ron (2 July 2014). "Luminoso Lands $6.5M In Series A To Keep Building Cloud Text Analytics Service" . TechCrunch. Retrieved 3 March 2015.

12. Darrow, Barb (11 February 2015). "Luminoso brings its text analysis smarts to streaming data" . GigaOm. Retrieved 3 March 2015.

13. Commonsense Computing Initiative (2009-02-24). "Divisi in Launchpad" . Retrieved 2009-10-07.

14. "The ConceptNet Project V2.1" . Retrieved 2008-12-17.

External links

- Open Mind Common Sense

- Open Mind Common Sense meta-repository Github

- ConceptNet

- AnalogySpace

- The Divisi inference toolkit

- Commonsense Computing Initiative's Webpage　　(Site doesn't exist)

- The Open Mind Initiative　　(Site doesn't exist)

- OMCSNetCPP - Open source C++ inference engine using the OMCSNet data

- Open Mind Common Sense in Brazil　　(Site broken)

- Advanced Interaction Laboratory