TOWARDS AN INTELLIGENT QUESTION-ANSWERING SYSTEM
State-of-the-art in the Artificial Mind

Andrej Gardoň and Aleš Horák
Faculty of Informatics, Masaryk University, Brno, Czech Republic

Keywords: Artificial Intelligence, Watson, True Knowledge, AURA, Dolphin-Nick, Transparent Intensional Logic, GuessME!, Mind Theory, Mind Axioms.

Abstract: This paper discusses three up-to-date Artificial Intelligence (AI) projects focusing on the question-answering problem – Watson, Aura and True Knowledge. Besides a quick introduction to the architecture of systems, we show examples revealing their shortages. The goal of the discussion is the necessity of a module that acquires knowledge in a meaningful way and isolation of the Mind from natural language. We introduce an idea of the GuessME! system that, by a playing simple game, deepens its own knowledge and brings new light to the question-answering problem.

1 INTRODUCTION

The idea of a machine, at least as intelligent as the human, has attracted many researches in the last few decades (Crevier, 1993; Goertzel & Pennachin, 2007; Hall, 2007). Generally, three sub-problems have to be solved: Data acquisition, data manipulation and data processing. While data manipulation is well-formed today, intelligent processing and data acquisition is far from the capabilities of our brains. Many tools for data storage are available, but there are few for ingenious information retrieval, especially those with natural language support. Such technologies, besides traditional data manipulation, provide data categorization, understanding the meaning and a deep question-answering mechanism. In this article we discuss three projects aimed at intelligent data processing. We analyse pitfalls of the mentioned systems and describe the project GuessME! which introduces the idea of automatic knowledge acquisition.

2 WATSON, AURA, TRUE KNOWLEDGE

A few months ago, the world was fascinated by an AI system called Watson introduced by IBM. As a competitor of the Jeopardy! quiz, it won the game and superseded the DeepBlue system (Hsu, 2002) in the chart of intelligent computers that defeated the most successful human players. It had no internet connection, no human interaction and was able to answer enough questions to win $77 147, leaving rivals at $24 000 and $21 600. Should we worry about our intellect? Definitely not! Although Watson is effective in factual problems, its abilities in creative tasks are limited. Let us quickly look at the system's architecture, internal processes and, by analysing some questions, reveal weaknesses of the system (based on Ferrucci, 2010 and YouTube archives of the Jeopardy! show).

The knowledge library is an essential component of a question-answering system like Watson. Millions of texts in different forms serve this purpose. Besides an unstructured approach (similar to Google), there is also a structured knowledge base (KB), storing entities and relations between them. IBM's research revealed the necessity to combine both methods. Usually, however, the KB must be provided in advance and most of the knowledge is stored in the unstructured form. These facts make Watson a nerd. He knows a lot, but he does not understand it.

Question-answering starts with a classification of questions and identification of sub-queries. Decomposed parts enter a phase of hypothesis generation and candidate answers are proposed by a

Gardoň A. and Horák A..
TOWARDS AN INTELLIGENT QUESTION-ANSWERING SYSTEM - State-of-the-art in the Artificial Mind.
DOI: 10.5220/0003831504060412
In *Proceedings of the 4th International Conference on Agents and Artificial Intelligence* (ICAART-2012), pages 406-412
ISBN: 978-989-8425-95-9

variety of search techniques (text/document/passage searches, KB querying, constraint lists). The amount of generated hypotheses was stabilized at 250 with a precision level of 85% (85% is the probability of generating a correct answer within the top 250 candidate answers for every question). Soft filtering based on lightweight scoring algorithms prune the initial set of answers and then proving by evidence begins. One of the most effective proving methods is a passage search. Here, a candidate answer is added to the original question's context and snippets of text satisfying both are retrieved. Finally, scoring and ranking algorithms identify the best answer. The system has noteworthy architecture that combines current data mining technologies and smart statistic methods for achieving the best results. But do human beings search tons of texts when answering a simple question? The Jeopardy! show is fast (usually 3-6 seconds per answer) and there is no time to conduct exhausting searches through one's knowledge. The most recent research identified synchronized patterns in frequencies of firing neurons. The highest frequencies represent an overall perception of an object while lower frequencies codify different visual aspects, emotions, etc. (Lane, 2009). Therefore, it is likely for the brain to store information in structured associations rather than pure texts. In school, one can try to be a nerd, but a clever teacher can always ask the question that reveals the true level of your understanding. In Watson, this is represented by a question from Game 1:

"From the Latin for "end", this is where trains can also originate".

Watson top three answers*:*

1. *finis (97%)*
2. *Constantinople (13%)*
3. *Pig Latin (10%)*

He had chosen the answer *"finis"* which was wrong. There is not a problem to infer the correct answer from the partial solution (*terminus, finis*) for a human. Another example comes from the *Name the decade* category in Game 1. Watson was not able to answer any questions from this category. He had the highest confidence in the first question:

Disneyland opens & the peace symbol is created

1. *1950s (87%)*
2. *Kingdom (6%)*
3. *It's a Small World (4%)*

however, he was superseded by a rival. The system most likely fails during a phase of evidence proving. It looks for sources meeting the requirements from both candidate answer and the question. Humans rather solve sub-queries and then join them Sources on Google related to the question

Disneyland opens & the peace symbol is created

1. *1920s (57%)*
2. *1910 (30%)*

support this theory. Watson preferred the answer *1920s* to the correct *1910s* (There are many sources containing key words from the question and 1910s).

The human brain's intelligence and limits of the Watson technology are revealed in the *Actors who Direct* category from Game 2. Human competitors recalled the answers while Watson had still been proving his hypotheses. However, other sections showed the advantages of Watson's methods. The strength of associations in the human brain determines the amount of knowledge and the level of reasoning used during the search for an answer. Therefore, some questions can take more time than that which is required by Watson's supercomputer. Brightness of human intellect overcomes this handicap in a brilliant way. A player with low confidence in an answer immediately buzzes in and takes five private seconds to seek the correct answer.

The *Also on your computer keys* category proves Watson's intelligence level. None of the proposed answers met the *computer key* constraint:

A loose-fitting dress hanging straight from the shoulders to below the waist

1. *chemise (97%)It's an abbreviation for Grand Prix auto racing*

1. *gpc (57%)*

The main disadvantage of Watson is the ignorance of the natural language (NL) meaning. A different approach can be found in the AURA project (prepared by Gunning, 2010), which attempts to pass advanced placement exams by learning from college-level science textbooks. During the development, three areas of interest were chosen (Biology, Physics and Chemistry) with selected sections in the textbooks. A trained expert in each domain was required to model the knowledge extracted from these texts. These responsible persons underlined the most important words in a paragraph. The highlighted sections were then mapped on concepts either by semantic search against a specialized knowledge base (SKB) or manually by the expert. Knowledge extraction was finished in a graph-editing tool where a concept map was created.

Besides textual entries, AURA can process tables and mathematical equations; however, diagrams and complex processes (as is the case in Biology) must be omitted. System querying is carried out in a simplified form of English:

A car is driving. The initial speed of the car is 12m/s. The final speed of the car is 25 m/s. The duration of the drive is 6.0 s. What is the distance of the drive?

Tests showed that AURA can correctly answer more than 70% of questions that were available to the experts during the creation of the SKBs (thus, it was possible to formulate the knowledge in a way that can easily reveal answers). When novel questions were asked, best results were achieved in Biology (47%), the worst in Chemistry (18%), which was caused by optimizing the SKBs to prior questions. The need for a trained expert to model all knowledge in AURA limits the system's usability. It would be more appropriate if the expert just supervised the learning process and answered potential questions formulated by the system. An inference module limits AURA in using built-in rules. As it is not possible to obtain new rules from NL, only a predefined set of problems can be solved.

True Knowledge (TK) is a project supporting automatic acquisition of knowledge from various sources (prepared by Tunstall-Pedoe, 2010). Relational databases can be mapped to TK format by specialized tools; summary tables found at the end of Wikipedia articles provide a structured informational resource; language processors extract data from unstructured parts of Wikipedia and Internet users can manually enter new knowledge. Each English sentence is simplified into

subject-noun phrase ↔ verb-phrase ↔
↔ object-noun-phrase

format, which is close to the one used by facts in KB (named relations between named entities). Besides simple facts, the KB can also have facts about facts and facts about properties of facts, all of which has the power to express many phenomena captured by NL. Consistency of the system is ensured by the inference mechanism that proposes the truthfulness of facts and rejects data causing contradictions. Inference rules are formed by generators programmed by people; this limits TK in the automatic creation of new rules.

Sentence analysis constrains the domain of acceptable problems. Each question is mapped on a template transforming NL into KB format. In case it is not possible to match a question with a template already present in the system, answer inferring fails. The following questions demonstrate the pitfalls of such a solution:

Who is the director of Rocky II? Sylvester Stallone
Who is the director of Rocky III? Sylvester Stallone
Who is the director of Rocky II and III? Fail

The system produces the best answers in simple factual questions (e.g. *"Who is Barrack Obama?"*), but an internal benchmark (by True Knowledge) showed only 17% of common questions can be answered. Although another 36% can be answered by adding new knowledge and a further 20% by creating new templates, poor results reveal the abilities of the self-learning system.

3 MIND MODULE

The discussed projects can be used in everyday life, but each of them lacks the intellect of the human brain. AURA and TK understand a portion of NL meaning, while Watson has great power to defeat human players without knowing what the nature of the question is. We identify the main problem in the core of all systems – acquisition of knowledge. Children require many years of studies to form an integrated view of the world. By games, books, problem-solving, they strengthen associations, tune concepts and create new reasoning rules. From childhood, human beings try to understand the outside world. It is, therefore, necessary to research a project that is able to learn in the same way as children. In this way, the system can remember the word "*apple",* with appropriate references to the real object, and further ask questions like: "*What is the colour of the apple? Is it food? Is the Apple a member of any class?"*

Natural language seems to be an essential component of intelligence but, as Steven Pinker says, it is rather an instinct (Pinker, 2000). Its main purpose is the communication of internal thoughts and awareness of external circumstances. In comparison to the senses (vision, hearing), it is rapid with effective exchange of information. However, the logic behind it is, according to the modular theory of Jerry Fodor (Fodor, 1983), likely joined with a separate module – the Mind. Two arguments support this proposition. First, the frontal lobe of the brain is identified as a centre of the Consciousness (Carter, 2009); the brain can process information from the senses, but one is not aware of it until this centre is activated. Thanks to this setup, we can walk along a familiar street and think something completely different. Secondly, learning by heart allows the reproduction of text without knowing what it is about (personally, I wonder about poems I have learned and never known about the meaning). Therefore, there are no doubts that handicapped (blind, deaf-mute) people can achieve high education levels even if they do not have a functioning channel of communication. Deaf-blindness is a loss of vision accompanied by lack of hearing, so the development of everyday language is excluded. Special communication methods based on touch are sufficient for those people to learn Mathematics (Řezáčová. 2007). As a conclusion, natural language is just another form of information coding with mediated reference of reality or the abstract world. We suppose there is a special module (call it the Mind), which supervises associations between different codes (sounds, pictures, words, etc.), providing inference capabilities and data processing. The Mind, with the cooperation of the Emotional module, forms a significant part of our intelligence. Recent research has revealed that all information from our senses meet in the Amygdala part of the brain (Carter, 2009) which is responsible for emotional reactions. If, let us say, that the connection between the vision and emotional centre is broken (as in Capgras' syndrome), you can clearly recognize the face of a familiar person, but you consider the person is a cheater as no appropriate emotion is invoked (Berson, 1983).

Despite the importance of the Emotional module, let us focus on the Mind, as it is essential for understanding coded information. Senses and NL have five common properties (CP). They can:

- Distinguish energetic fields called Objects (Apple, Car, Red colour, Singing …);
- Identify properties and parts of objects (red, cold, leg …) that are themselves objects;
- Describe relations between objects (a man has a leg, a man has a father...);
- Analyse the dynamics of objects (I ate an apple); and
- Categorize objects into concepts to provide general properties of its members.

Grammatical categories in the sentence *"Smart Watson won the Jeopardy! game."* express some CP. *Watson* and *Jeopardy! game* are objects, *smart* is a property of *Watson* and the verb *won* describes an activity performed by *Watson* (dynamics of an object). You can realize CP by senses with a simple test. Close your eyes and take an ice cube into your hand. You inspect it as a sole object that is cold and melts in time. Formal logic systems usually lack some aspect of CP (e.g. first-order logic is unable to represent the dynamics of objects) and, therefore, their computational equivalents cannot reach the required level of intelligence.

Transparent Intensional Logic (TIL) represents NL meaning in an algorithmically accessible form and fully supports CP (Tichý, 2004). It is designed to analyse all information from the sentence

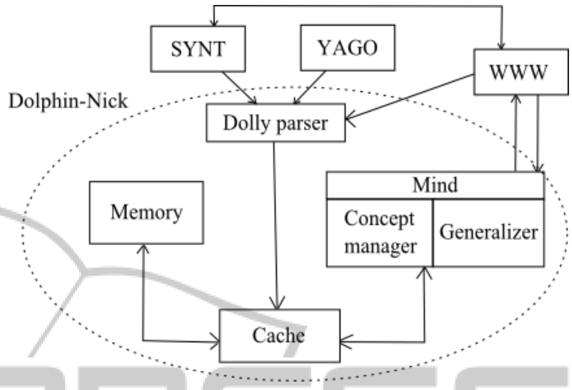

Figure 1: GuessME! architecture.

(temporal aspect, personal attitudes, beliefs, etc.) and code it in the form of a construction. A sentence with its corresponding representation in TIL follows:

Andrej was shopping in the supermarket on (this) Friday.

$$\lambda w\, \lambda t\, [\mathbf{P}_t\, [\mathbf{Thr}_w\, \lambda w_1\, \lambda t_1\, [\mathbf{Does}_{w_1 t_1}\, Andrej\, [\mathbf{Perf}_{w_1}\, \mathbf{to\_shop\_in\_supermarket}_{w_1}]]]\, \mathbf{Friday}]$$

Joining TIL with a question-answering module (QAM) is the idea of the GuessME! system.

4 GUESSME! SYSTEM

GuessME! is a system based on a simple game for two players. One player chooses an Object (see the definition above) or Event and the other one has to guess this object by asking questions. Actions and relations are excluded from the possible domain; however, questions can contain these actions ("*Is it used for washing?"*). There are two operational modes:

- Game - the user chooses whether he/she will guess or think and then questions are postulated to reveal the object
- Explorer – the computer asks about objects from the KB to form new associations or to confirm the truthfulness of previous knowledge

By asking questions about data already present in the system, GuessME! is able to deepen knowledge associations, generalize information, form concepts or even create new inference rules. It also extracts the meaning of NL and stores it in an internal format (Dolly Construction, DC; see Gardoň, 2010). Comparing this to the 20-Questions game (Speer et al., 2009), GuessME! is an open domain, supports typed NL and is two-way (humans can be the guesser).

Figure 1 shows the architecture of the GuessME! system. The computational equivalent of TIL called Dolphin-Nick (Gardoň & Horák, 2011) is used as a KB. This system is capable of processing TIL constructions, supports the temporal aspect (Gardoň & Horák, 2011) and allows basic forms of inference. A brief introduction of modules follows (For more information consult Gardoň, 2011):

SYNT is a tool for automatic transcription of NL sentences to corresponding TIL constructions (Horák, 2008). It provides an NL language interface to the Dolphin-Nick KB.

WWW stands for *Why?What?Where?* and represents the QAM module. It is responsible for generating questions and answers. Besides simple *Yes/No* questions, it is possible to ask a question having a set of simple words as an answer (e.g. "*What is the colour of X, What classes is X member of?").*

YAGO is based on a KB containing more than 10 million entities and 80 million facts about them (Suchanek, Kasneci, Weikum, 2008). It is used to collect common knowledge and alternatively to get additional information about previously stored data. In GuessME!, it is possible to enter information like *"a car is a thing"* and the system uses YAGO to obtain further information.

DOLLY parser is a tool for converting TIL constructions into DC. As a DC is language independent, GuessME! can be adapted to any language.

CACHE is a temporary storage place for incoming information (see Gardoň & Horák, 2011).

MEMORY is organized as a semantic network of DCs.

MIND manages inference rules denoted by sentences like *"Every man is human."* The internal mechanism checks the consistence of the KB using these rules. GENERALIZER can automatically create new rules from a probability table (PT) defined by a concept. Every set (TIL object of type *(ο)ξ*) in the Dolphin-Nick system corresponds to a Concept with representative individual (RI) sharing properties of all set members. CONCEPT MANAGER creates PT according to proportional coverage of properties (see Figure 2) and GENERALIZER takes top rows with 100% coverage to make new rules from them. The dynamic nature of such rules is clear.

One of the TIL advantages is a theory of possible worlds (Tichý, 2004) – the Dolphin-Nick KB can contain knowledge with different truthfulness depending on possible worlds used (*The world is flat* can be *true* in the KB itself but *false* in a world describing a model connected with the user *Peter*). Worlds are used to model personal attitudes and play GuessME!. When a game starts, a new individual is created in the game world (GW). With progress, answers are transformed into a model represented by the GW, which is continually checked against the general KB world to propose new questions. When there is enough confidence in the character of the guessed individual, the system tries to guess its name.

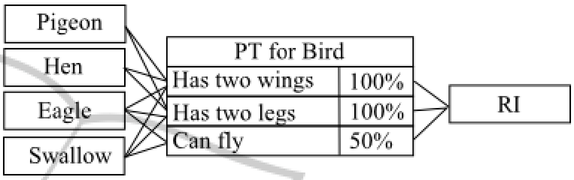

Figure 2: Concept for the word Bird.

At the beginning of a game, players agree on the type of object being guessed (Object or Event). In the case of an Event, temporal questions can be used with full support of time tenses (see Gardoň & Horák, 2011), e.g. thinking of the day America was discovered, one can ask *"Did this event happened during last millennium? Was it before or after Christ?, etc."*.

The GuessME! project is under development and we are intensively working on its modules. It is necessary to provide an interface connecting YAGO with Dolphin-Nick and examine methods for acquisition of knowledge from this KB. The Mind is partially implemented with basic inference rules. The temporal aspect is also fully supported. Further steps are focused on the CONCEPT MANAGER, GENERALIZER and a complex inference module (especially on the capability of identifying rules in a text and their incorporation into the Mind). Strategy of game play is to be devised and formulation of questions must be specified within the WWW module. The key step is to formulate common knowledge, which allows the playing of the first games. School textbooks from the first grades of elementary education will be used to teach the system basic facts.

We hope that GuessME!, by simulation of human progress through education, will lead to a complex question-answering machine.

Table 1: Summary of discussed question-answering systems.

	Watson	AURA	True Knowledge	GuessME!
Type of knowledge	Mostly unstructured	Structured	Structured	Structured
Input method	Encyclopedias, DBs, texts	Logic formalism	DBs, Wikipedia, Users	Users, YAGO, School textbooks
Question formulation	NL	Simplified NL	NL templates	NL
Pros	Unrestricted domain	Can solve mathematical problems	Automatic acquisition of knowledge	Unrestricted domain, full NL support, acquisition of knowledge
Cons	Does not really understand NL	Input method, domain specific	Unable to answer complex questions	-
Usability	Data mining	Education tool	New Google	New Google, smart Wikipedia

5 CONCLUSIONS

In this article we have discussed three different projects in Artificial Intelligence that have a common goal – the question-answering issue. We identified their shortfalls and proposed intelligent acquisition of knowledge as a solution. An overview of presented systems is summarized in Table 1. The GuessME! System, based on a simple game, is introduced as a basic step towards a Watson-like system with full NL support. It combines structured knowledge in the form of a KB (like AURA), natural language as the main communication method (True Knowledge, Watson), open-domain orientation (Watson, True Knowledge) and a theory of possible worlds. The nature of the GuessME! project uncovers our mistrust in systems like Watson. As a human being must undergo years of studies to become an intellectual adult, the same must be done within a computer system. GuessME! should be the first step.

ACKNOWLEDGEMENTS

This work has been partly supported by the Czech Science Foundation under the project P401/10/0792.

REFERENCES

Berson, R. J., 1983. Capgras' syndrome. *American Journal of Psychiatry*, 140(8), pp.969-978.

Carter, R., Aldridge, S., Page, M., & Parker, S. 2009. The Human Brain Book. London: DK ADULT.

Daniel Crevier. 1993. Ai: The Tumultuous History of the Search for Artificial Intelligence. *New York: Basic Books*.

Ferrucci, D. et al. 2010. Building Watson : An Overview of the DeepQA Project. *AI Magazine* 31(3): p.59-79.

Fodor, J. 1983. The Modularity of Mind. Massachusetts: *The MIT Press*.

Gardoň, A., 2010. Dotazování s časovými informacemi nad znalostmi v transparentní intenzionální logice (Queries with temporal information over knowledge in Transparent Intensional Logic), in Slovak, Master thesis, *Brno: Masaryk university, Faculty of informatics.*

Gardoň, A., 2011. The Dolphin Nick project. [online] Available at: <http://www.dolphin-nick.com>.

Gardoň, A., & Horák, A. Knowledge Base for Transparent Intensional Logic and Its Use in Automated Daily News Retrieval and Answering Machine. In D. T. Steve (ed), *3rd International Conference on Machine Learning and Computing*. Singapore 26-28 February 2011. Chengdu, China: IEEE.

Gardoň, A., & Horák, A. 2011. Time Dimension in the Dolphin Nick Knowledge Base Using Transparent Intensional Logic. In I. Habernal & V. Matoušek (eds), *Proceedings of the 14th international conference on Text, speech and dialogue*. Plzeň, Czech Republic 1-5 September 2011. Berlin, Heidelberg: Springer-Verlag

Goertzel, B., & Pennachin, C. 2007. Artificial general intelligence Ben Goertzel & Cassio Pennachin (eds). *Springer*.

Gunning, D. et al. 2010. Project Halo Update-Progress Toward Digital Aristotle. AI Magazine 31(3): p.33-58.

Hall, J. S. 2007. *Beyond AI: creating the conscience of the machine. New York: Prometheus Books*.

Horák, A., 2008. Computer Processing of Czech Syntax and Semantics. *Brno, Czech Republic: Librix.eu*.

Hsu, F.-hsiung. 2002. Behind deep blue: building the computer that defeated the world chess champion. *New Jersey: Princeton University Press*.

Lane, N. 2009. Life ascending : the ten great inventions of evolution. *New York: W.W. Norton*.

Pinker, S. 2000. The Language Instinct : How the Mind Creates Language (Perennial Classics). *New York: Harper Perennial Modern Classics*.

Řezáčová, P., 2007. Specifika edukačního procesu u jedinců s hluchoslepotou (Particularities of education process for deaf-blind people), in Czech, Master thesis, *Brno: Masaryk university, Faculty of education*.

Speer, R. et al. 2009. An interface for targeted collection of common sense knowledge using a mixture model. In C. Conati & M. Bauer (eds), *Proceedings of the 14th international conference on Intelligent user interfaces*. Florida, USA 8-11 February 2009. New York, USA: ACM.

Suchanek, F. M., Kasneci, G., & Weikum, G. 2008. YAGO: A Large Ontology from Wikipedia and WordNet. *Elsevier Journal of Web Semantics* 6(3): p.203-217.

Tichý, P., 2004. Collected Papers in Logic and Philosophy. Prague: Filosofia, *Czech Academy of Sciences, and Dunedin: University of Otago Press*.

Tunstall-Pedoe, W. 2010. True Knowledge: Open-Domain Question Answering Using Structured Knowledge and Inference. *AI Magazine* 31(3).