SEMANTIC NETWORK AS A KNOWLEDGE BASE IN TRAINING SYSTEMS

Igor A. Bessmertnyy, Viacheslav S. Kulagin
Saint Petersburg State University of Information Technologies, Mechanics and Optics
Saint Petersburg, Russia
Igor_bessmertny@hotmail.com

Training system (TS) is a subclass of expert systems, i.e. knowledge based information systems. Knowledge representations in TS may differ depending on their purpose and the topic. Simple TS implement computerized questionnaires and the knowledge base may consist of questions and answers. TS based on expert system shells use facts and rules, like all the expert systems do. The paper presented describes an approach to TS building where knowledge presented by semantic network.

Semantic network as a knowledge base has a lot of advantages: syntax and language independence, visibility, simplicity of inserting new information. The item of semantic network is a triple including a subject, an object and the predicate that defines the relation between them. Additional dictionaries help to eliminate problem of synonymy and polysemy and let arrange a simple natural language interface because the program has to match just a user three-word sentence with a triple in the semantic network.

The Program "Semantic", the first release of TS based on these principles, written in Visual Prolog, supports a verbal human-computer dialog and displays the growing graph of facts established during the conversation. The program is included to educational process in St.Petersburg State University of Informational Technologies, Mechanics and Optics.

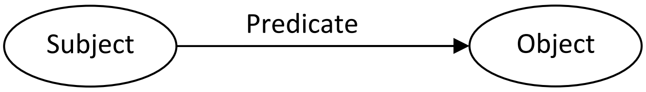
The paper describes a class of training systems where the dialog initiative belongs to the trainee. That means the human asks and the computer answers. The computer simulates a studied object and the trainee interacts with it like with a real one. For instance, these training systems may simulate the doctor-patient interaction. The training system chooses a hypothesis and the user should identify that asking questions. The quality of trainee’s knowledge and skill is a minimal number of questions asked before diagnose is determined.

To build such kind of dialog the training system needs a knowledge base. The simplest knowledge base is a list of expected questions and answers of the machine like this (let say we try to diagnose the car trouble):

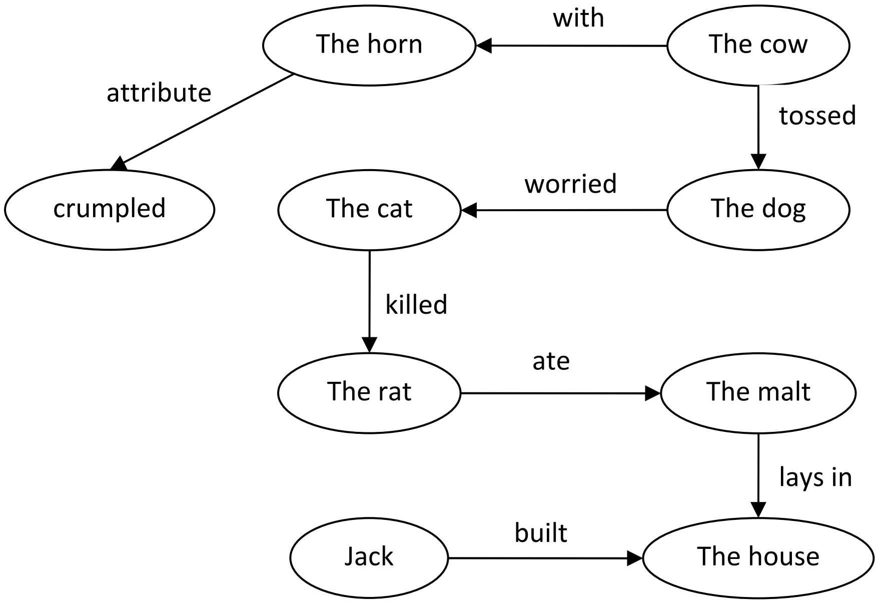
What is a car brand?	The car brand is Toyota.
What is a problem with the car?	The problem is the car’s engine.
Does the car’s engine start?	The car’s engine does not start.

This knowledge base is very easy to create but it is hard to build an inference engine. Identification of questions is a serious problem because any question may be asked in many varieties. A compromise solution is identification of questions by keywords (Bessmertny, Kulagin, 2006). For the example above the keywords might be “car” “brand”, “car” “problem” and “engine” “start”. This method allows recognize simple user’s phrases pretty reliable if the knowledge base is not large.

A prospective form of knowledge presentation is semantic network. An item of semantic network is a triple,



where predicate means relation between subject and object. A set of linked triples build the network describing the topic. The main advantage of semantic network is lingual independence, i.e. abstraction from language syntax. Each term may be expressed in any language including pictograms. The picture below contains a fragment of the well known verse “The house that Jack built”.



Simple data structure of triples yields simple queries to knowledge base. A complex sentence separates to a set of triples, which can be easily recognized. Very simple parser with syntax analyzer identifies sentences in affirmative and interrogative form. An example of knowledge base in Prolog syntax, presented in (Bessmertny, 2007), is listed below. Hereinafter Prolog predicate “f” means a fact, “d” – a dictionary, “j” – junk words.

```
f(“car”,  “has_a”, “brand”).   % Facts
f(“car”,  “has_a”, “model”).
f(“car”,  “has_a”, “year”).
f(“brand”,  “is”, “Toyota”).
f(“model”,  “is”, “Camry”).
f(“year”,  “is”, “2007”).
f(“car”,  “has_a”, “engine”).
f(“engine”,  “does_not”, “start”).
```

The simple data structure of triples yields simple queries to the knowledge base in a natural language. To recognize questions like

Is the brand Toyota?
What is the year?
Does the engine start?

we have to skip articles and then recognize question words if any and identify the sentence model: with or without a question word. To do this the knowledge base should be completed by appropriate Prolog predicates:

```
q([“what”, “where”, “when”, “why”, “who”, “how”]).        % Question words
j([“a”, “an”, “the”, “to”]).                               % Articles and junk words
d([“do”, “does”, “does_not”, “did”]).                     % Synonyms
```

Thus, we can store facts into the knowledge base and extract them.

Knowledge representation by semantic networks has two main problems: synonymy and polysemy. Synonymy means several terms for one entity. This problem could be solved easily using a dictionary of synonyms. By the way, dictionaries allow build multilingual semantic networks, like this:

```
d([“car”,  “PKW”, “voiture”]).
```

Polysemy means number of meanings of the same term. Polysemy is much wider problem than it seems at the first glance at words like “field”. The word “field” has a dozen of meanings, but “Jack” that built the house – several millions. Two thousand years ago “Jesus from Nazareth” identified a unique person. Nowadays only Social Security Code or Taxpayer Individual Code can distinguish a person. However, in everyday interactions we eliminate misunderstanding without strict objects identification because of context.

Context is information environment that implicitly completes every phrase to give a sense. Context control is very important part of dialog. At the start of dialog context is low; each spoken sentence replenishes the context. By the way, the verse “The house that Jack built” is an example of text, where context does not raise. In contrary, the text itself grows.

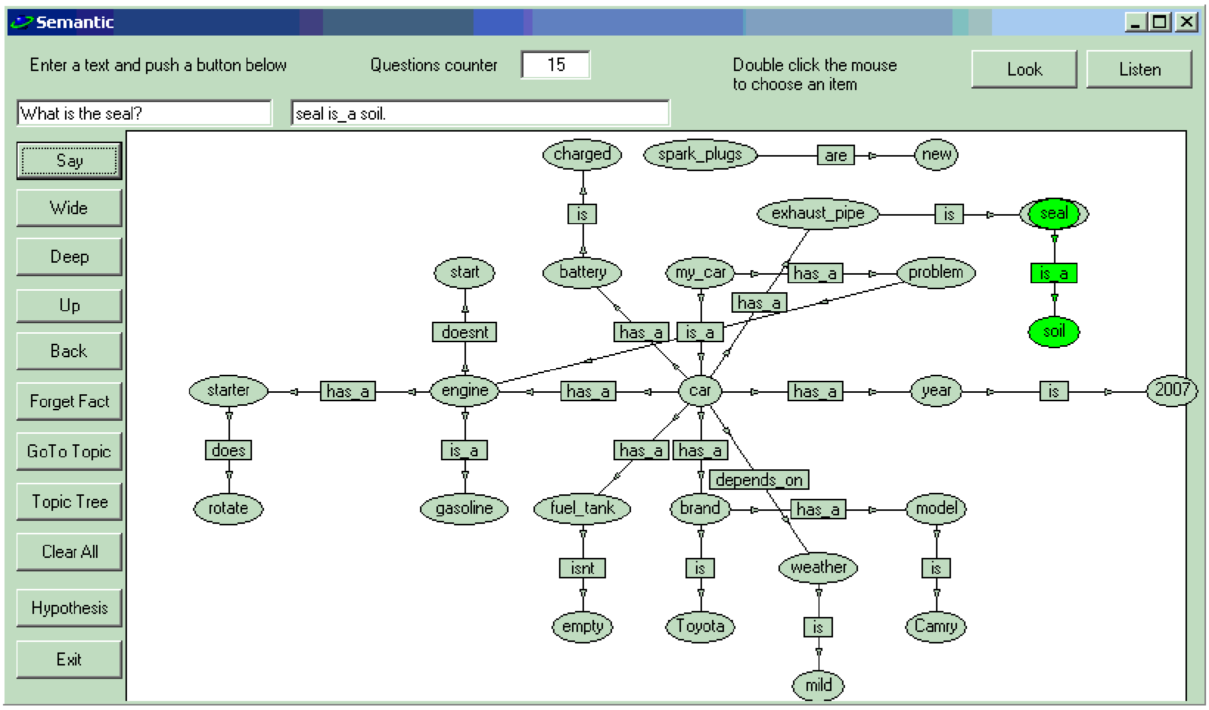
An implementation of training system is written in Visual Prolog 7.0. The user interface lets display the dialog development as a growing graph of semantic network. This allows to reminder to the user the dialog context. To avoid the polysemy problem the current topic must not be too large. The program allows switch between topics, i.e. save the current context and load another one. This happens by a real conversation when we say: “Now let’s talk about X”. This key phrase implies we have to forget the current context and remember all previously discussed facts about the topic X.

Now let’s try to apply the program to find out a car trouble. The knowledge base should contain a separate file for each diagnosis and the program chooses a hypothesis i.e. the file randomly so the trainee should identify the trouble cause. Below is a dialog that leads successfully to the goal.

```
What is a brand?
- brand is Toyota .
What is a model?
- model is Camry .
What is a year?
- year is 2007 .
What is an engine?
- engine is_a gasoline .
What is a problem?
- problem is_a engine .
Does the engine start?
- engine doesnt start .
Is the battery full?
- battery is charged .
Is the fuel_tank empty?
- fuel_tank isnt empty .
Does the starter rotate?
- starter does rotate .
Are spark_plugs ok?
- spark_plugs are new .
Is the weather cold?
- weather is mild .
Is the exaust_pipe ok?
(014E) Error unification stmt "Is exaust_pipe ok ?".
Is the exhaust_pipe ok?
- exhaust_pipe is sealed .
What is the seal?
- seal is_a soil .
Selected topic is "sealed".
- seal is_a soil .
- You are right. This is the true diagnose.
```

Actually, when the driver reversed the car, it bumped into a heap of ground and the exhaust pipe has blocked. That’s why the engine does not start. The statement with a grammar mistake is left here intentionally to demonstrate what happens if the user misprints.

Beyond the text dialog the program supports drawing the graph of semantic network that helps to see the knowledge base. The screenshot below shows the user interface features in general and the semantic network of the studied topic. Let us remark that the program does not draw entire network graph but just facts already discovered by the trainee. Otherwise the user gets a clue to find the solution.



Semantic

Enter a text and push a button below

Questions counter 15

Double click the mouse to choose an item

What is the seal?

seal is_a soil.

Thus, the implementation of a training system has shown that semantic network is a good knowledge media for training and information systems that allows easily as create and modify knowledge bases as build a training system shell. The first release of program “Semantic” used at the Computer Science Department of Saint Petersburg State University of Informational Technologies, Mechanics and Optics in the learning courses “Artificial Intelligence” and “Computer Technologies in Science”.

References

I. Bessmertny, V. Kulagin (2006). An approach to building learning expert systems. Proceedings of the 10th IACEE World Conference on Continuing Engineering Education, Vienna, Austria, 2006.
http://www.wccee2006.org/papers/343.pdf

I. Bessmertny (2007). An Intellectual Agent in Training Systems. Proceedings of the 5th International Conference on Education and Information Systems, Technologies and Applications: EISTA 2007, Orlando, Fl, USA, 2007.

Curriculum Vitae

Igor A. Bessmertnyy: Associate Professor (since 1992), Computer Science Department, Saint Petersburg State University of Information Technologies, Mechanics and Optics. Received EE from Riga Civil Aviation Institute (1976), Riga, Latvia, and PhD from Leningrad Institute of Fine Mechanics and Optics (1986). Senior Scientist at the Leningrad Institute of Fine Mechanics and Optics (1984-1992).

Viacheslav S. Kulagin: Dean of a Faculty (since 1976), Professor at Electronic Department (since 2002), Saint Petersburg State University of Information Technologies, Mechanics and Optics. Received CS from Leningrad Institute of Fine Mechanics and Optics (1968). Senior Scientist in 1973. Department Individual Member IACEE since 001 and ASCEE since 2003. Member emeritus of the Russian Federation Higher School staff.