Semantic Analysis and Question Answering: a System Under Development

Igor Boguslavsky, Vyacheslav Dikonov, Leonid Iomdin, Alexander Lazursky, Victor Sizov, Svetlana Timoshenko

A. A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow
The General Layout of the Question Answering System

- Output: a table containing an element (or elements) from the repository of individuals.
- Question processing is divided into three stages:
  1) Syntactic analysis;
  2) Semantic analysis;
  3) Creation of a SPARQL Query.
The Repository of Individuals

- contains information on individual objects and situations that are concrete instances of Ontology concepts.
- Individual objects: Moscow, France, the Thames, Cervantes.

It can be enriched manually or semi-automatically with the help of a program gathering information from external databases like dbpedia, but the main (and inexhaustible!) source is texts and their semantic representations. Thus, semantic analysis is the most important part of our system: it is involved both in question processing and in collecting data for the repository.
Semantic analysis

• Performed by the ETAP-3 module;
• Produces a semantic structure, which consists of a set of triples of the type

\texttt{semantic\_relation (individual,individual)}

and can be seen as an RDF graph. Semantic relations and individuals constituting this structure correspond to elements of the ontology.
Semantic analysis: Starting from Syntactic structure

• Уже на второй минуте матча вратарь гостей достал мяч из ворот.
‘As early as on the second minute, the goalkeeper took the ball out of the goal’

Dialogue 2015, Moscow
Semantic Analysis:
Step 1: Basic Semantic Structure

<table>
<thead>
<tr>
<th>Гипотезы 14</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2 TimeInterval</td>
<td>2.7 NUMERAL(2)</td>
<td></td>
</tr>
<tr>
<td>4.1 FootballMatch(*)</td>
<td>10.1 Minute(*)</td>
<td></td>
</tr>
<tr>
<td>6.1 SportsTeam</td>
<td>12.1 Region</td>
<td></td>
</tr>
<tr>
<td>5.1 Human</td>
<td>6.1 SportsTeam</td>
<td></td>
</tr>
<tr>
<td>11.1 GoalkeeperRole(*)</td>
<td>13.1 Region</td>
<td></td>
</tr>
<tr>
<td>6.1 SportsTeam</td>
<td>3.2 TimeInterval</td>
<td></td>
</tr>
<tr>
<td>5.1 Human</td>
<td>8.2 BallForSports(*)</td>
<td></td>
</tr>
<tr>
<td>12.1 Region</td>
<td>9.5 GoalArea(*)</td>
<td></td>
</tr>
<tr>
<td>12.1 Region</td>
<td>13.1 Region</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Фраза</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2 УЖЕ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.7 NUMERAL(2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dialogue 2015, Moscow
Semantic Analysis: Representation of Word Meanings

Вратарь ‘goalkeeper’ – human who occupies the goalkeeper position.

hasRole (Human_1, GoalkeeperRole_1)

Гости ‘guests’ – the team that trains on a stadium different from that of the current match.

represents (SportsTeam_1, Region_1)
hasLocation (FootballMatch_1, Region_2)
hasParticipant (FootballMatch_1, SportsTeam_1)
differentFrom (Region_1, Region_2)
Semantic Analysis: Inference Rules

• (i) if a physical object is taken out of some place, it was previously located in this place (the precondition of the Taking event)

• (ii) if a ball is located in the goal area, this is the result of the goal (the effect of the GoalEvent; this rule is a kind of abductive inference)
Semantic Analysis
Step 2: Applying Inference Rules

Dialogue 2015, Moscow
Two types of Semantic Decomposition

In the example above we could see two types of semantic decomposition:

1. Russian-semantic rules that map word meanings to the elements of ontology. They build a graph if there is no one-to-one correspondence (semantic representations of words ВРАТАРЬ ‘goalkeeper’, ГОСТИ ‘guest team’),

2. Intrasemantic rules that interpret concepts (Precondition rule for the Taking concept)
The role of Ontology

• a source of structured world knowledge and tool for creating the repository of individuals
• a metalanguage of semantic representations.

Our ultimate goal is to provide all (non-functional) Russian words with an interpretation in terms of ontological semantic elements.

OntoEtap ontology has two sources:
• upper/ middle ontology SUMO (Suggested Upper Merged Ontology, http://www.adampease.org/OP/) and
• our own small ontology of the football news domain, written in OWL. OntoEtap is maintained in Protégé environment.

Today OntoEtap contains 7,438 classes and 3,757 individuals.
Beyond the Sentence Boundaries

• Basic hypothesis:
Considering the fact that news messages normally (though not always) cover only one sports event, we assume that different mentions of an event within a message are coreferent if no evidence to the contrary are available. Pointing to different times, locations, or lists of participants may form such evidence.
Beyond the Sentence Boundaries: an example

Source message:
• Мюнхенская "Бавария" обыграла мадридский "Реал" в первом полуфинальном матче футбольной Лиги чемпионов. Встреча, проходившая на стадионе "Альянц-Арена" в Мюнхене, завершилась со счетом 2:1.
‘Munich Bavaria beat Real Madrid in the first semifinal match of the football Champions League. The meeting, held at the stadium “Allianz Arena” in Munich, ended with the score 2:1’

Question:
• Как сыграли «Бавария» и «Реал Мадрид»?
‘How did Bavaria and Real Madrid play?’
The structure could be read as follows: “The team SportsTeam_1 is called Bayern (which is German for Bavaria) and represents Munich (lines 1–2). It lost to team SportsTeam_2 in a semifinal match SemifinalMatch_1 (lines 3–5). Team SportsTeam_2 is called FC-Real-Madrid and represents Madrid (line 6–7). The semifinal match in which these teams participated is part of the Champions League and has sequence number 1 (lines 8–12)."
Beyond the Sentence Boundaries: an example

Dialogue 2015, Moscow
Beyond the Sentence Boundaries: OWL Document
Contribution of the Repository of Individuals

If the user asks
Какая команда нанесла поражение чемпиону Испании?
‘What team defeated the champion of Spain?’
while we have a text reporting that
Футболисты "Барселона" обыграли мадридский "Атлетико“, ‘Barcelona’s footballers beat "Atletico“ (Madrid)’
then the system will establish the correspondence between the question, the text, and the records in the repository of individuals, and will come with the correct answer.
Contribution of the Repository of Individuals

Semantic structure of the question.

Dialogue 2015, Moscow
Contribution of the Repository of Individuals

Semantic structure of the sentence.

Dialogue 2015, Moscow
Contribution of the Repository of Individuals

Background knowledge on Atletico Madrid

Dialogue 2015, Moscow
The answer to the question: Whom did «Bavaria» beat?