

Ю. М. Смирнов, А. М. Андреев,
Д. В. Березкин, А. В. Брик

ВЕРОЯТНОСТНЫЙ СИНТАКСИЧЕСКИЙ АНАЛИЗАТОР ДЛЯ ИНФОРМАЦИОННО- ПОИСКОВЫХ СИСТЕМ

Исследованы современные подходы к задаче синтаксического анализа естественного языка — одной из наиболее сложных проблем искусственного интеллекта. Вероятностные анализаторы обладают меньшей трудоемкостью разработки, не требуют высокой лингвистической подготовки, при этом обладают достаточно высоким качеством работы, которое в большинстве случаев может быть еще улучшено путем дополнительного обучения. В вероятностных алгоритмах синтаксического анализа используются метод дерева принятия решений, различные статистические представления и модели грамматики, эвристические методы оптимизации и т.д. Дан вывод о том, что комбинированный подход, сочетающий как формальные описания грамматики, так и вероятностные механизмы анализа, представляется более эффективным.

Probabilistic syntactic analyser for information retrieval systems / Yu.M. Smirnov, A.M. Andreev, D.V. Berezkin, A.V. Brik // Vestnik MGTU. Priborostroenie. 2000. No. 2. P. 34–54.

State-of-the-art approaches to the problem of syntactic analysis of natural language are considered. This subject is one of the most complicated problems in the artificial intelligence domain. Probabilistic analysers are characterised by lower development laboriousness, do not need high linguistic training; they operation is qualitative enough, and in most cases it can be improved by additional training. In the probabilistic algorithms of syntactic analysis the tree method of making decisions, various statistic presentations and grammar models, heuristic optimisation methods, etc., are used. The conclusion is made that the combined approach uniting both formal grammar descriptions and probabilistic analysis mechanisms, seams to be most efficient. Tabs.1. Refs.13.

СПИСОК ЛИТЕРАТУРЫ

1. Смирнов Ю. М., Андреев А. М., Березкин Д. В., Брик А. В. Об одном способе построения синтаксического анализатора текстов на естественном языке // Изв. вузов. Приборостроение. – 1997. – Т. 40. – № 5. – С. 34–42.
2. Андреев А. М., Березкин Д. В., Брик А. В. Лингвистический процессор для информационно-поисковой системы // Компьютерная хроника. – 1998. – № 11. – С. 79–100.

3. Андреев А. М., Березкин Д. В., Брик А. В. Кантонистов Ю. А. Вероятностный синтаксический анализатор для информационно-поисковой системы // Компьютерная хроника. – 1999. – № 1. – С. 37–85.
4. Manning C. D., Carpenter R. Probabilistic Parsing Using Left Corner Language Models. 1997 // Опубликовано на сервере www.xxx.lang.gov/cmp-ig.
5. Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. Т. 1. Синтаксический анализ. – М.: Мир, 1978. – 612 с.
6. Фу К. Структурные методы в распознавании образов. – М.: Мир, 1977. – 312 с.
7. Losee E. M. Learning Syntactic Rules and Tags with Genetic Algorithms for Information Retrieval and Filtering: An Empirical Basis for Grammatical Rules // Information Processing & Management, 1995 // Опубликовано на сервере www.xxx.lang.gov/cmp.lg.
8. Magerman D. M. Natural Language Parsing as Statistical Pattern Recognition // A dissertation submitted to the department of computer science at the committee on graduate studies of Stanford University, 1994 // Опубликовано на сервере www.xxx.lang.gov/cmp.lg.
9. Интеллектуальные системы автоматизированного проектирования БИС и СБИС / В.А. Мищенко, Л.М. Городецкий, Л.И. Гурский и др.: Под ред. В.А. Мищенко. – М.: Радио и связь, 1988. – 272 с. – С. 109–125.
10. Андреев А. М., Березкин Д. В., Кантонистов Ю. А. Экспертные юридические системы: миф или реальность // Мир ПК. – 1998. – № 9. – С. 56–64.
11. Карпова Г. Д., Пирогова Ю. К., Кобзарева Т. Ю., Микаэлян Е. В. Компьютерный синтаксический анализ: описание моделей и направлений разработок // Итоги науки и техники Сер. “Вычислительные науки”. – Т. 6. – М.: ВИНТИ, 1991.
12. Искусственный интеллект. В 3-х кн. Кн. 2. Модели и методы: Справочник / Под ред. Д.А. Поспелова. – М.: Наука, 1990. – 304 с.
13. Попов Э. В. Общение с ЭВМ на естественном языке. – М.: Наука, 1982. – 360 с.

Статья поступила в редакцию 30.03.2000

Юрий Матвеевич Смирнов родился в 1923 г., окончил МВТУ им. Н.Э. Баумана. Д-р техн. наук, профессор кафедры “Компьютерные системы и сети” МГТУ им. Н.Э. Баумана. Лауреат Государственной премии СССР. Автор более 200 научных работ в области вычислительных средств и систем управления.

Yu. M. Smirnov (b. 1923) graduated from the Bauman Moscow Higher Technical School. D. Sc. (Eng.), professor of “Computer Systems and Networks” Department of the Bauman Moscow State Technical University. USSR State Prize winner. Author of more than 200 publications in the field of computation means and control systems.

Арк Михайлович Андреев родился в 1943 г., окончил в 1967 г. МВТУ им. Н.Э. Баумана. Канд. техн. наук, доцент кафедры “Компьютерные системы и сети” МГТУ им. Н.Э. Баумана. Имеет более 70 научных работ в области вычислительных средств и систем управления.

A.M. Andreev (b. 1943) graduated from Moscow Higher Technical School in 1967. Ph. D. (Eng.), ass. professor of “Computer Systems and Networks” Department of the Bauman Moscow State Technical University. Author of more than 70 publications in the field of computation means and control systems.

Дмитрий Валерьевич Березкин родился в 1966 г., окончил в 1990 г. МГТУ им. Н.Э. Баумана. Канд. техн. наук, директор Научно-производственного центра “Интелтек Плюс”. Имеет около 30 научных работ в области вычислительных средств.

D.V. Berezkin (b. 1966) graduated from the Bauman Moscow State Technical University in 1990. Ph. D. (Eng.), director of Scientific-Industrial Center “INTELTEK PLUS”. Author of more than 30 publications in the field of computer technique.

Алексей Владимирович Брик родился в 1974 г., окончил в 1997 г. МГТУ им. Н.Э. Баумана. Аспирант кафедры “Компьютерные системы и сети” МГТУ им. Н.Э. Баумана. Специализируется в области использования формальных грамматик и методов синтаксического анализа в информационных технологиях.

A.V. Brik (b. 1974) graduated from the Bauman Moscow State Technical University in 1997. Post-graduate of “Computer Systems and Networks” Department of the Bauman Moscow State Technical University. Specialises in application of formal grammar and syntactic analysis methods in information technique.