

DOI: 10.18454/2079-6641-2017-18-2-76-80

ИНФОРМАЦИОННЫЕ И ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ
УДК 004

РАЗРАБОТКА МАТЕМАТИЧЕСКОЙ МОДЕЛИ АВТОМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ ЗНАНИЯ ДЛЯ ГИБРИДНОЙ ВОПРОСНО-ОТВЕТНОЙ СИСТЕМЫ

А. С. Вальчук

Камчатский государственный университет имени Витуса Беринга, 683032,
г. Петропавловск-Камчатский, ул. Пограничная, 4
E-mail: valchukos@kamgu.ru

В работе предложена и реализована, в виде математической модели и программного комплекса, концепция использования коэффициентов ценности, для ранжирования ответов в универсальных гибридных вопросно-ответных системах, которая позволит повысить эффективность их работы.

Ключевые слова: алгоритм ранжирования ответов, гибридные вопросно-ответные системы, коэффициенты ценности ответов

© Вальчук А. С., 2017

INFORMATION AND COMPUTER TECHNOLOGIES
MSC 68U35

DEVELOPMENT OF A MATHEMATICAL MODEL OF AUTOMATIC KNOWLEDGE EXTRACTION FOR A HYBRID QUESTION-ANSWER SYSTEM

A. S. Valchuk

Vitus Bering Kamchatka State University, 683031, Petropavlovsk-Kamchatsky,
Pogranichnaya st., 4, Russia
E-mail: valchukos@kamgu.ru

The paper proposes and implements, in the form of a mathematical model and software package, the concept of using value coefficients, for ranging answers in universal hybrid question-answer systems, which will improve the efficiency of their work.

Keywords: algorithm for ranging answers, hybrid question-answer systems, coefficient of value of answers.

© Valchuk A. S., 2017

Введение

В процессе поиска ответа на вопросы в интернете человеку приходится отсеивать большие объемы лишней информации [1,2]. Автоматические интеллектуальные вопросно-ответные системы, стараются отвечать на естественном языке, но ещё далеки от совершенства [3,4,5]. Анализ же наиболее известных гибридных вопросно-ответных систем (ГВОС) [6], к которым можно отнести форумы [8] и различные виды интернет-чатов показал, что практически все они или не имеют каких-либо серьезных механизмов ранжирования и дополнительной обработки ответов, или слишком сложны и специализированны и не подходят для использования их алгоритмов в универсальных системах [7]. Обычно пользователям просто пересылаются ответы других участников в соответствии с датой их поступления в систему. Наиболее ценный и полезный ответ от наиболее компетентного специалиста может оказаться в самом низу длинного списка и остаться незамеченным. Для решения этой проблемы и улучшения качества работы ГВОС в нашей работе предлагается использовать специальный алгоритм ранжирования ответов. Он сможет учитывать, как мнение большинства по каждому вопросу, так и компетентность каждого отдельного пользователя и автоматически находить наилучший баланс между ними.

Постановка задачи и метод решения

Для решения обозначенной проблемы была разработана концепция коэффициентов ценности ответов. То есть список ответов предлагается ранжировать в соответствии с коэффициентом ценности вычисляемым для каждого из них. Это позволит легче и быстрее находить пользователям нужную и более качественную информацию. Целью работы стала разработка математической модели автоматического извлечения знания, и программы ГВОС, реализующих идею использования коэффициентов ценности ответов. Такая программа позволит оценить и оптимизировать целевую функцию, используя данные о её работе в реальных условиях.

В упрощенном виде коэффициент ценности Kp ответа (i) вычисляется по формуле:

$$Kp(i) = \frac{\sum Reit(i)}{Rel(i)}, Rel(i) = \frac{\sum Lev(i)(j)}{N(i)},$$

где $\sum Reit(i)$ - сумма рейтингов (уровней) автора ответа по категориям (i) соответствующим данному вопросу (рейтинг пользователя напрямую зависит от общего качества генерируемых им ответов, то есть от их оценок другими людьми), $Rel(i)$ - коэффициент релевантности ответа, $N(i)$ - количество ответов по вопросу, $\sum Lev(i)(j)$ - сумма расстояний Левенштейна ответа (i) до остальных ответов (j) по данному вопросу.

Расстояние Левенштейна [9] определяется как минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую. Этот коэффициент показывает насколько похожи между собой два ответа, а их сумма дает представление о том насколько конкретный ответ (i) похож на все остальные. Насколько он релевантен некому "общему мнению" пользователей, ответивших на вопрос (i).

В процессе практического использования системы ценность той или иной составляющей коэффициента будет меняться. Для того, чтобы иметь возможность в

Kw- показатель качества работы системы

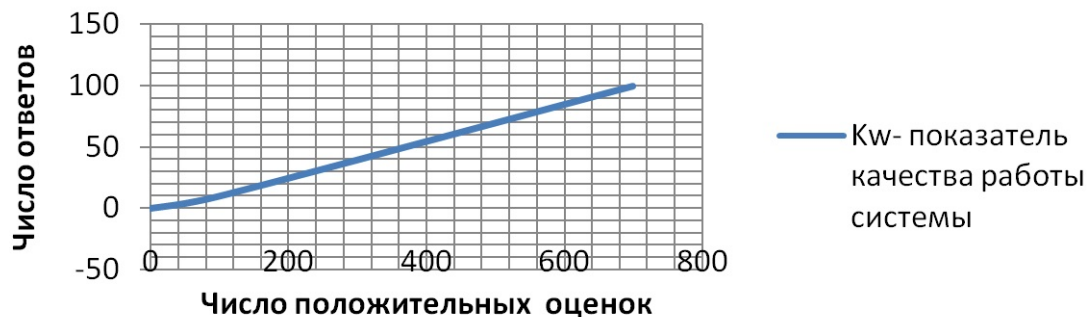


Рисунок. График Kw. Показатель эффективности системы.

соответствии с практическими данными подстраивать систему, (для ее более эффективной работы) в алгоритм были введены два регулятора:

$T(reit)$ - произвольная величина (периодически изменяющая значение (в автоматическом режиме или по указанию оператора) для корректировки ценности рейтинга пользователя;

$T(rel)$ - произвольная величина для корректировки ценности релевантности ответа.

Получаем:

$$Kp(i) = \frac{\sum Reit(k) * T(reit)}{Rel(i) * T(rel)}$$

Для оценки качества работы системы, был введен новый индикатор (показатель) Kw. По динамике его изменения можно судить о необходимости корректировки системы или о результатах воздействия на неё. Например - изменения регулирующих коэффициентов, числа пользователей, и т.д..

$$Kw \approx \frac{N(stars)}{N(o)},$$

где $N(0)$ - общее число ответов по всем вопросам на данный момент, $N(stars)$ - общее число оценок (звезд) по всем вопросам на данный момент.

Заключение

В результате проделанной, на данный момент, работы были выделены объекты, предметы и гипотезы исследования. Разработана целевая функция для нахождения коэффициентов ценности и ранжирования ответов. Написана программа ГВОС (web-приложение) [10,11], реализующая целевую функцию ранжирования ответов и проведены её тестовые испытания с реальными пользователями для проверки и улучшения работы целевой функции.

Список литературы

- [1] Шокин Ю. И., Барахнин В. Б., Федотов А. М., *Проблемы поиска информации*, Новосибирск: Наука, Новосибирск, 2010, 134 с. [Shokin Ju. I., Barahnin V. B., Fedotov A. M. Problemy poiska informacii. Novosibirsk: Nauka, 2010. 134].
- [2] Воскресенский А. Л., Хахалин Г. К., “Формирование запросов к поисковой машине для извлечения знаний из Интернета”, *Компьютерная лингвистика и интеллектуальные технологии*, Тр. Междунар. конференции "Диалог 2005", Наука, М., 2005, 86-91. [Voskresenskij A. L., Nahalin G. K. Formirovanie zaprosov k poiskovoj mashine dlja izvlechenija znanij iz Interneta. Komp'juternaja lingvistika i intellektual'nye tehnologii. Tr. Mezhdunar. konferencii "Dialog 2005". Moskva. Nauka, 2005. 86-91.].
- [3] Гаврилова Т. А., Хорошевский В. Ф., *Базы знаний интеллектуальных систем*, 2000. [Gavrilova T. A., Horoshevskij V. F. Bazy znanij intellektual'nyh sistem, 2000.].
- [4] Тихомиров И. А., “Вопросно-ответный поиск в интеллектуальной поисковой системе Exactus”, *Труды четвертого российского семинара по оценке методов информационного поиска РОМИП*, 2006, 80-85.. [Tihomirov I. A. Voprosno-otvetnyj poisk v intellektual'noj poiskovoj sisteme Exactus. Trudy chetvertogo rossijskogo seminaru po ocenke metodov informacionnogo poiska ROMIP. 2006. 80-85.].
- [5] Соловьёв А. А., Пескова О. В., “Построение вопросно-ответной системы для русского языка: модуль анализа вопросов”, *Новые информационные технологии в автоматизированных системах*, **13** (2010). [Solov'jov A. A., Peskova O. V. Postroenie voprosno-otvetnoj sistemy dlja russkogo jazyka: modul' analiza voprosov. Novye informacionnye tehnologii v avtomatizirovannyh sistemah. 2010. vol. 13].
- [6] Венда В. Ф., *Системы гибридного интеллекта: Эволюция, психология, информатика*, 1990. [Venda V. F. Sistemy gibridnogo intellekta: Jevoljucija, psihologija, informatika. 1990].
- [7] Колесников А. В., Кириков И. А., Листопад С. В., *Гибридные интеллектуальные системы с самоорганизацией: координация, согласованность, спор*, ИПИ РАН, М., 2014. [Kolesnikov A. V., Kirikov I. A., Listopad S. V. Gibridnye intellektual'nye sistemy s samoorganizaciej: koordinacija, soglasovannost', spor. Moskva. IPI RAN, 2014.].
- [8] Селютин А. А., “Форум как жанр виртуального общения”, *Вестник Челябинского государственного университета*, **9** (2008). [Seljutin A. A. Forum kak zhanr virtual'nogo obshhenija. Vestnik Cheljabinskogo gosudarstvennogo universiteta. 2008. vol. 9.].
- [9] Левенштейн В. И., “Двоичные коды с исправлением выпадений, вставок и замещений символов”, *Доклады Академий Наук СССР*, 1965, 845-848. [Levenshtejn V. I. Dvoichnye kody s ispravleniem vupadenij, vstavok i zameshhenij simvolov. Doklady Akademij Nauk SSSR. 1965. 845-848].
- [10] Бенедетти Р., *Изучаем работу с jQuery*, Издательский дом "Питер", 2012. [Benedetti R. Izuchaem rabotu s jQuery. Spb.: Izdatel'skij dom "Piter 2012.].
- [11] Нестеренко С. О., *Технологии создания веб-сайтов*, 2013. [Nesterenko S. O. Tehnologii sozdanija veb-sajtov. 2013.].

Список литературы (ГОСТ)

- [1] Шокин Ю.И., Барахнин В.Б., Федотов А.М. Проблемы поиска информации. Новосибирск: Наука, 2010. 134 с.
- [2] Воскресенский А. Л., Хахалин Г. К. Формирование запросов к поисковой машине для извлечения знаний из Интернета. Компьютерная лингвистика и интеллектуальные технологии // Тр. Междунар. конференции "Диалог 2005". М.: Наука, 2005. С. 86-91.
- [3] Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем, 2000.
- [4] Тихомиров И. А. Вопросно-ответный поиск в интеллектуальной поисковой системе Exactus // Труды четвертого российского семинара по оценке методов информационного поиска РОМИП. 2006. С. 80-85.
- [5] Соловьёв А. А., Пескова О. В. Построение вопросно-ответной системы для русского языка: модуль анализа вопросов // Новые информационные технологии в автоматизированных системах. 2010. Т. 13.

- [6] Венда В. Ф. Системы гибридного интеллекта: Эволюция, психология, информатика. 1990
- [7] Колесников А. В., Кириков И. А., Листопад С. В. Гибридные интеллектуальные системы с самоорганизацией: координация, согласованность, спор. М.: ИПИ РАН, 2014.
- [8] Селютин А. А. Форум как жанр виртуального общения // Вестник Челябинского государственного университета. 2008. Т. 9.
- [9] Левенштейн В. И. Двоичные коды с исправлением выпадений, вставок и замещений символов // Доклады Академий Наук СССР. 1965. С. 845-848
- [10] Бенедетти Р. Изучаем работу с jQuery. Спб.: Издательский дом "Питер" 2012.
- [11] Нестеренко С. О. Технологии создания веб-сайтов. 2013.

Для цитирования: Вальчук А. С. Разработка математической модели для гибридной вопросно-ответной системы // *Вестник КРАУНЦ. Физ.-мат. науки*. 2017. № 2(18). С. 76-80. DOI: 10.18454/2079-6641-2017-18-2-76-80

For citation: Valchuk A. S. Development of a mathematical model for a hybrid question-answer system, *Vestnik KRAUNC. Fiz.-mat. nauki*. 2017, **18**: 2, 76-80. DOI: 10.18454/2079-6641-2017-18-2-76-80

Поступила в редакцию / Original article submitted: 07.06.2017