

Стороженко А.В. Особенности проведения социологического опроса // Академия педагогических идей «Новация». Серия: Студенческий научный вестник. – 2018. – №11 (ноябрь). – АРТ 549-эл. – 0,2 п.л. - URL: <http://akademnova.ru/page/875550>

РУБРИКА: ТЕХНИЧЕСКИЕ НАУКИ

УДК 681.5

Стороженко Алексей Владимирович
магистрант 2 курса, факультет морской
Литвинов Александр Николаевич
магистрант 2 курса, факультет морской
ФГБОУ ВО «Керченский государственный морской
технологический университет»
г. Керчь, Российская Федерация
e-mail: esiap@mail.ru

ОСОБЕННОСТИ ПРОВЕДЕНИЯ СОЦИОЛОГИЧЕСКОГО ОПРОСА

Аннотация: Автоматизация информационных процессов в социально-экономических системах охарактеризована увеличением объема разнородной информации во разных отраслях человеческой деятельности и соответствующими потребностями пользователей оперативно, полно и качественно получать документные и электронные первоисточники.

Ключевые слова: данные, функция, параметры, структура, модель.

Storozhenko Alexey Vladimirovich
2nd year undergraduate, marine faculty
Litvinov Alexander Nikolaevich
2nd year undergraduate, marine faculty
FGBOU VO "Kerch State Marine Technological University"
Kerch, Russian Federation

FEATURES OF CONDUCTING A SOCIOLOGICAL SURVEY

Abstract: Automation of information processes in socio-economic systems is due to the continuously increasing amount of various information in all branches of human activity and the corresponding needs of users to quickly, fully and efficiently obtain documentary and electronic sources.

Keywords: data, function, parameters, structure, model.

При извлечении, обработке, сохранении информации возникают вопросы, связанные с формальным ее описанием. От степени формализации и представления информации зависит эффективность используемой информационной технологии, а также уровень автоматизации, определяемый степенью участия человека в процессе получения информации. Под информационной технологией будем понимать ее определение, данное академиком Глушковым В.М..

Чем точнее формальное представление информации, тем выше возможность ее компьютерной обработки в организационных системах. Можно выделить три вида информации, поступающей на обработку в информационную систему: формализованная, неформализованная и слабоформализованная информация. Под формализованной или структурированной информацией будем понимать данные. Данные представляют собой факты и характеризующие их числовые, количественные показатели: имена, даты событий, сведения об экономических процессах, местах действия, то есть это информация, обработанная специальным образом для последующего принятия решений [1-4].

Под слабоформализуемой информацией будем понимать разноформатные файлы, содержащие тексты на естественном языке, в которых сложно выделить факты, объекты, со своими свойствами и связями. Такая информация, содержащаяся в интранет- и Интернет-сетях, не может быть успешно обработана информационными системами и реализована в виде знаний при отсутствии эффективного доступа. Документы содержат слабоформализуемую информацию. Под неформализуемой информацией обычно понимается различного рода вербальная информация, в которой в своем большинстве используются вербально-коммуникативные сигналы и методы. Обычно, под вербальными сигналами понимаются словесные формулировки (разговор), тон голоса, его тембр, звуки и восклицания.

Под знаниями будем понимать информацию, необходимую для принятия решений, т.е. истинную, достоверную и проверенную практикой информацию [4]. Знания делят на факты и эвристики, глубинные и поверхностные, жесткие и мягкие. Факты указывают на хорошо известные обстоятельства. Эвристики основываются на индивидуальном опыте специалиста. Поверхностные знания представляют знания, полученные из опыта, внешние эмпирические ассоциации с каким-либо явлением.

Системы Text Mining применяются для: отнесения документов к некоторым критериям из заданной схемы их систематизации (classification); кластеризации (clustering); обеспечения поиска документов по ключевым словам (keyword searching); тематического индексирования (thematic indexing); извлечения фактов и понятий (feature extraction); разработки тезаурусов (thesauri); построения семантических сетей; ответов на запросы (question answering); автоматического реферирования и квазиреферирования (automatic text summarization).

Среди систем, реализующих большинство представленных технологий, наиболее известными являются: Intelligent Miner for Text (IBM), PolyAnalyst (Megaputer Intelligent), Text Miner (SAS), Semio Map (Semio Corp.), Oracl Text (Oracle), Knowledge Server (Autonomy), RetrievalWare (Convera).

Например, система Intelligent Miner for Text корпорации IBM представляет один из лучших инструментов глубинного анализа текстов и позволяет автоматически определять язык, на котором составлен документ, автоматически относить документ к некоторой категории, разбивать большое множество документов на группы по близости стиля, формы и различных частотных характеристик выявляемых ключевых слов, выделять в документе собственные имена и названия и определять их в качестве ключевых слов и т.п. Естественно, расстояние между образами $\alpha = \alpha(\omega_i, \omega_j)$ определяется как вероятность, так как вероятность является нормой. Таким образом, источник генерирует образы с определенными вероятностями, система идентификации выдвигает гипотезы из заданного алфавита эталонов с соответствующими вероятностями. Степень достоверности гипотезы определяется расстоянием между образом, реализованным источником, и порождаемой эталоном гипотезой. Структура системы в этом случае имеет вид процедуры проверки гипотез (рис. 1).

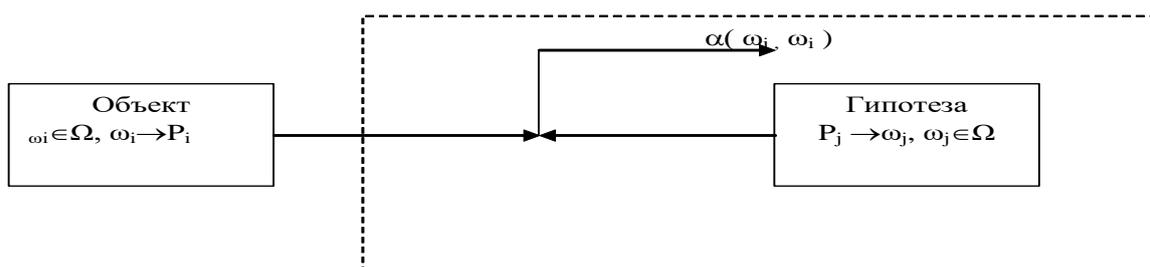


Рисунок 1. Структура системы с проверкой гипотезы.

Естественно, сохраняя формальное условие формирования метрики как нормы разности, в данном случае нормы разности множеств $\alpha(\omega_i, \omega_j) = \|\omega_i \setminus \omega_j\|$ следовательно, расстояние в пространстве образов формируется как вероятность разности множеств

$$\alpha(\omega_i, \omega_j) = P(\omega_i \setminus \omega_j), \quad (1)$$

так как справедливо

$$\omega_i \setminus \omega_j = \omega_i \bar{\omega}_j, \quad (2)$$

выражение (2) можно записать в виде

$$\alpha(\omega_i, \omega_j) = P(\omega_i \bar{\omega}_j). \quad (3)$$

Событие ω_i всегда рассматривается как входное, поэтому учитывая, что в силу полноты ансамблей справедливо

$$P(\bar{\omega}_i / \omega_j) = 1 - P(\omega_i / \omega_j) \quad (4)$$

и исключив дополнение, получаем метрику пространства образов в виде

$$P(\bar{\omega}_i \omega_j) = P(\omega_j)(1 - P(\omega_i / \omega_j)). \quad (5)$$

В данном случае целесообразно нормировать полученное отношение на вероятность входного образа, что позволит оценивать расстояние между образами ω_i и ω_j условной вероятностью.

Проверим выполнение условий для метрики

$$\alpha(\omega_i, \omega_j) = 0 \Leftrightarrow \omega_i = \omega_j;$$

$$\alpha(\omega_i, \omega_j) \geq 0;$$

$$\alpha(\omega_i, \omega_k) \leq \alpha(\omega_i, \omega_j) + \alpha(\omega_j, \omega_k),$$

Следствием этого является разнообразие способов структурирования данных в странице, так что два логически однородных элемента данных могут быть форматированы совершенно различным образом, либо даже содержать ошибки, т.е. части кода могут не вполне соответствовать синтаксису иерархической структуры тегов.

Список использованной литературы:

1. Mizzaro S. How many relevances in information retrieval?, in C.W. Johnson and M. Dunlop (eds) / S. Mizzaro // Proceeding of the Workshop «Information Retrieval and Human Computer Interaction», GIST Technical Report GR96-2 / Glasgow University. — Glasgow: The British Computer Society— P. 57-60.
2. Nagao M. A framework of a mechanical translation between Japanese and English by analogy principle, in Artificial and Human Intelligence / M. Nagao; ed. A. Elithorn, R. Banerji. — North Holland, 1984. — P. 173 – 180.
3. O'Brien I.A. Management Information Systems: A Managerial End User Perspective / I.A. O'Brien — Boston: IRVIN, 1990.— 650p.
4. W3C Semantic Web Activity. — Режим доступа: <http://www.w3.org/2001/sw/>.
5. Wiig K. Knowledge management is no illusion! / K. Wiig // Proc. of the First International Conference on Practical Aspects of Knowledge Management. – Zurich: Swiss Informaticians Society, 1996.

Дата поступления в редакцию: 28.11.2018 г.

Опубликовано: 29.11.2018 г.

© Академия педагогических идей «Новация». Серия «Студенческий научный вестник», электронный журнал, 2018

© Стороженко А.В., 2018