

УДК 519.68

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА РАСПОЗНАВАНИЯ ТЕКСТА СТРОИТЕЛЬНОЙ ТЕМАТИКИ

Ю.Н. Тесля¹, Е.Е. Шабала²

¹ профессор, ² ассистент.

¹ Киевский национальный университет имени Тараса Шевченко,

² Киевский национальный университет строительства и архитектуры (Киев), Украина

***Аннотация.** Рассматривается подход к применению элементов искусственного интеллекта для разработки интеллектуальной информационной системы формирования, ведения и актуализации ресурсной нормативной базы в строительстве. Одной из основных целей есть распознавание текста строительной тематики.*

***Ключевые слова:** нормативная база, интеллектуальная система, искусственный интеллект, система прогнозирования, строительный процесс, фрагмент текста, база знаний*

Постановка проблемы.

Современное состояние и особенности строительства, такие как продолжительность эксплуатации строительных объектов, его роль в экономическом развитии государства обуславливают необходимость и актуальность совершенствования нормативной базы в этой области. Нормативная база влияет на соблюдение требований надежности и безопасности строительных объектов и регламентацию проведения проверок соответствия продукции этим требованиям. Все это требует повышения эффективности информационных технологий, обеспечивающих решение задач управления строительством на основе гибкой интеллектуальной информационной системы формирования и ведения ресурсной нормативной базы. Решение этих проблем возможно только с применением в разрабатываемых информационных технологиях современных методов и средств прогнозирования, подходов к созданию интеллектуальных систем формализации содержания произвольных текстов. Любое строительство описывается значительными по объему текстами (текстовыми документами), которые необходимо структурировать, классифицировать, формализовать. Поэтому эффективность информационных технологий в этой отрасли во многом зависит от языковых средств общения с компьютером. Отечественные и зарубежные системы обработки естественно – языковой информации носят либо экспериментальный характер, или имеют узкую специализацию и не удовлетворяют требованиям информационных технологий. Кроме того, и расходы на создание таких систем значительны, что не позволяет их широко использовать [замінити на сучасну книгу]. Но, тем не менее, без таких средств значительно снижается производительность автоматизированных систем при наполнении информационной базы, повышаются расходы на выполнение рутинных действий, таких как кодирование информации, поиск в информационной базе и др.[2].

Анализ исследований.

Внимание исследователей в этой области с самого начала было ориентировано на выявление закономерностей, организующих смысловое единство текста. На первом этапе этих работ наиболее популярными были подходы, основанные на выявлении тех или иных статистических закономерностей распределения терминов в тексте или их взаимного расположения в нем [3, 4]. В дальнейшем исследования в области автоматизированного реферирования сместились в сторону использования внутренних структур текста, выявление той информационной основы, которая организует весь текст [5, 6]. Также возникла необходимость учитывать внешние воздействия, изменяющие информационное содержимое базы знаний. Создание фундаментальной теории, обобщающей и формально описывающей закономерности в процессах взаимодействия в Природе на основе гипотезы об их пезиловом характере, формально представляющую активную и объективную роль информации в этих процессах позволило по-новому подойти к созданию интеллектуальных информационных систем.

Формулировка цели статьи.

Целью статьи является демонстрация преимуществ рефлекторной интеллектуальной системы обработки естественно-языковых текстов в строительной отрасли для совершенствования управления процессом формирования, ведения и актуализации ресурсной нормативной базы в строительстве.

Изложение основного материала.

Интеллектуальные системы, разрабатываемые в рамках интрофизического подхода, получили название рефлекторных, поскольку, по аналогии с высшей нервной деятельностью живых существ, умеют формировать ответные реакции на различные комбинации входных воздействий. Под рефлексом интеллектуальной системы будем понимать ответную реакцию интеллектуальной системы на входные воздействия, вырабатываемую специальными алгоритмом, который называется рефлекторным [2]. Для разработки рефлекторной интеллектуальной системы автоматизации процесса формализации содержания произвольных текстов строительной тематики (в дальнейшем РИС), целесообразно объединить требования к различным подсистемам с общими требованиями. Функциональными требованиями к разработке рефлекторной системы формализации содержания произвольных текстов являются:

- информационная система должна иметь интуитивно понятный интерфейс для удобной работы пользователя;
- трудоёмкость обработки данных не должна превышать квадратичную;

- система должна поддерживать сохранение семантики данных при их импорте и экспорте.

К основным нефункциональным требованиям относятся:

- информационная система должна быть разработана в соответствии с действующими правилами и стандартами на разработку программного обеспечения;
- отсутствие необходимости менять программный код при изменении структуры данных.

Данные требования будут в дальнейшем использованы при проектировании и реализации модулей интеллектуально-информационной системы. Требования к разработке РИС можно удовлетворить благодаря использованию рефлекторного алгоритма обработки естественно-языковых текстов. Критериями качества функционирования РИС являются:

- развитая коммуникативная способность;
- способность решать разнопрофильные задачи;
- способность к самообучению;
- способность к адаптации.

Для обеспечения удовлетворительности этих критериев и как следствие практической реализации результатов научных исследований была разработана структура базы данных рефлекторной системы формализации содержания произвольных текстов, алгоритм и программные средства этой системы. Система разработана в среде MS Access с использованием языка программирования VBA Access. Рефлекторная система формализации смысла произвольных текстов строительной тематики состоит из следующих блоков:

- Добавление предложений в базу данных – позволяет добавить название текста и сам текст.
- База данных - создана для хранения предложений и ключевых понятий.
- Добавление ключевых понятий в базу данных – позволяет ввести новые ключевые понятия или выбрать из существующего списка ключевые понятия.
- База данных ключевых понятий – создана для хранения названий ключевых понятий из строительной отрасли.

Для решения задачи автоматизации формализации смысла произвольных текстов строительной тематики авторами разработана экспериментальная система определения ключевых строительных понятий в естественно-языковом тексте.

Этапы осуществления формализации содержания произвольных текстов:

1. Добавление предложений в базу данных.
2. Добавление содержания в предложения.
3. Выбор типа расчета (авто расчет, расчет с ограничениями).
4. Ввод параметров расчета (количество новых фрагментов за один цикл обучения, максимальная/минимальная длина фрагмента, границы длины фрагмента, минимальное число статистики).
5. Выбор режима перебора предложения (по словам, по буквам).
6. Выбор типа перебора (по всем, по верхним уровням).
7. Осуществление нормализации текста.
8. Расчет и вывод статистики.

Входными данными являются предложение и список ключевых понятий. Предложение отражается от процессов строительства. Ключевое понятие – это то, о чем идет речь в тексте. Пример предложений и ключевых понятий из обучающей выборки приведен на рис. 1.

Назва тексту		КНАУФ				
Добавити	База	Обучение				
Назва тексту	Г	навчальна вибірка	Зміст 1	Зміст 2	Зміст 3	Зміст 4
3914	КНАУФ	Оценка материалов				
Они делают профиль жестким и удобным для строителей при монтаже конструкций						
3915	КНАУФ	технология производства раф	Профиль			
Для установки подвеса с разжиленным элементом края стенок профилей загнуты в середину и служат как упор						
3916	КНАУФ	технология производства раф				
Профили направляющие изготавливаются с готовыми перфорированными отверстиями по середине основы диаметром 6 мм через каждые 260 мм, предназначенными для установления дюб						
3917	КНАУФ	Оценка материалов	Гипс			
Гипсокартон отлично подходит для внутренней отделки стен						
3918	КНАУФ	Применение материала				
С его помощью можно сделать абсолютно ровные стены, в то время как просто штукатуркой добиться такого результата сможет лишь настоящий профессионал						
3919	КНАУФ	Применение материала				
Гипсокартон подходит для кирпичных, блочных и каркасных стен						
3920	КНАУФ	технология производства раф				
В статье рассмотрим как крепить гипсокартон при помощи алюминиевого профиля						
3921	КНАУФ	технология производства раф				
49% Профиль 47% технология производства раф 12% Картон 11% Применение материала 10% Пере						

Рис. 1. Фрагмент обучающей выборки

Алгоритм содержит 3 этапа:

1. Определяются абсолютные и условные вероятности появления ключевых понятий (при условии, что в исходном тексте есть некоторое буквосочетание).
2. Отклонение условной вероятности от абсолютной вероятности появления этого ключевого понятия свидетельствует о несилловом (информационном) влиянии интеллектуального аппарата эксперта, написавшего эти тексты на возникновение образа, соответствующего ключевому понятию.
3. По максимальной сумме таких воздействий выбираются ключевые понятия [2].

На Рис.2 приведены результат расчета статистики для определения содержания текста строительной тематики.

Статистика	
Всего	140
Угадано	87
Ложная тревога	357
Пропуск цели	53
Предложений	100
% угадано	62,14%
Средняя лож.тревога	3,57
Средний пропуск цели	0,53

Рис. 2. Результат расчета статистики определения смысла текста

Выводы.

В связи с большим объемом информации в строительной отрасли, которую нужно обрабатывать, необходимо создать такую систему, которая бы позволила рационально управлять этой информацией. Экспериментальные примеры показали высокий результат распознавания текста строительной тематики, которая была бы полезна для успешного поиска и структуризации нормативной базы в строительной отрасли. Поэтому пример разработки такой системы, приведенный в этой статье может стать основой для дальнейших исследований в этом направлении, который совместил бы в себе средства искусственного интеллекта, моделирования и прогнозирования выполнения строительных проектов.

СПИСОК ЛИТЕРАТУРЫ

1. Искусственный интеллект: в 3-х т. / под ред. Э. В. Попова, Д.А. Поспелова. – М. : Радио и связь, 1990 р.
2. Тесля, Ю. Н. Введение в информатику природы / Ю. Н. Тесля // Монография. – Киев : Кондор, 2010. – 256 с.
3. Михайлов, А. И. Основы информатики. / А. И. Михайлов, А. И. Черный, Р. С. Гиляревский. – М. : Наука, 1968.
4. Леонов, Б. П. О методах автоматического реферирования (США 1958–1974 гг.) / Б. П. Леонов // Научно-техническая информация, Сер. 2. – 1975. – № 6. – С. 16–20.
5. Пащенко, Н. А. Проблемы автоматизации индексирования и реферирования / Н. А. Пащенко, Л. В. Кнорина, Т. В. Молчанова, и др. // Итоги науки и техники. Сер. Информатика. – М. : ВИНТИ, 1983. – Т. 7. – С. 7–164.
6. Севбо, И. П. Структура связного текста и автоматизации реферирования. / И. П. Севбо. – М. : Наука, 1969.

Материал поступил редакцию 26.06.14.

INTELLIGENT SYSTEM OF TEXT RECOGNITION OF CONSTRUCTION SUBJECT

Yu.N. Teslya¹, E.E Shabala²

¹ Professor, ² Assistant

¹ Taras Shevchenko National University of Kyiv,

² Kyiv National University of Construction and Architecture (Kyiv), Ukraine

Abstract. The article contains the approach to the usage of elements of artificial intelligence for the development of the intellectual information system of formation, maintenance and updating of the resource regulatory frameworks in construction. One of the main goals is the text recognition of construction subjects.

Keywords: regulatory framework, intelligent system, artificial intelligence, forecasting system, construction process, a piece of text, the knowledgebase.