**Парамонов Антона Іванович. Інтерпретація текстової інформації на основі нечітких концептуальних моделей : Дис... канд. наук: 05.13.06 – 2009**

|  |  |
| --- | --- |
| |  | | --- | | **Парамонов А.І. Інтерпретація текстової інформації на основі нечітких концептуальних моделей. – Рукопис.**  Дисертація на здобуття наукового ступеня кандидата технічних наук зі спеціальності 05.13.06 – «Інформаційні технології». – Донецький національний університет, Донецьк, 2008.  У дисертації зроблено огляд сучасного стану проблеми автоматизації бізнес-процесів, виділено основні напрями її рішення та обґрунтовано її актуальність. Здійснено огляд сучасних інформаційно-пошукових систем, форматів представлення інформації в мережах і служб їх обробки. Наведено критерії оцінки результатів роботи пошукових систем. Зроблено огляд стану проблеми розуміння природно-мовних текстів. На підставі відомостей з когнітивної психології розроблено нечітку гібридну модель представлення знань. Розроблено метод інтерпретації текстової інформації на основі нечіткої концептуальної моделі. Введено нечітку характеристику мереж моделі – активність, на базі формалізованої когнітивної моделі логогену Мортона. Етап концептуального аналізу побудований на основі моделі активності мереж, згідно з якою смисл тексту формується з урахуванням трьох складових. Формалізовано механізм міркувань на основі досвіду у вигляді моделі інтерпретації тексту. Описано технологію побудови автоматизованих систем інтерпретації тексту та методика побудови баз знань. Запропоновано організацію, реалізацію та оцінку автоматизованих систем інтерпретації тексту. | |
| |  | | --- | | У дисертаційній роботі запропоновано нове рішення актуальної проблеми пошуку в інформаційних системах і набули подальший розвиток автоматизовані системи інтерпретації тексту. Основні результати роботи сформульовано таким чином.   1. Проведений огляд форматів представлення інформації в мережі Інтернет і додатків, які працюють з мережевими ресурсами, показав неузгодженість цих форматів, що значно ускладнює процес автоматичної обробки текстової інформації. Для вирішення цього питання запропоновано використовувати єдиний формат RSS, побудований на основі XML, що став фактично стандартом для зберігання, передачі і обробки динамічної інформації. 2. Проведений аналіз сучасних систем пошуку виявив наступні недоліки:   – індексний пошук, по-перше, не забезпечує достатнього рівня показників релевантності і пертинентності, по-друге, не дозволяє ефективно працювати з динамічним контентом, що ускладнює використання його для автоматизації бізнес-процесів, по-третє, велика роль в отриманні результатів покладається на користувача;  – існуючі інтелектуальні системи, по-перше, інтегруються з пошуковими системами за ключем, що в результаті знижує показник релевантності; по-друге, механізми самонавчання систем слабо автоматизуються, по-третє, підходи до інтелектуального пошуку не дозволяють інтегрувати різні аспекти проблеми розуміння ПМ текстів у єдину модель.  Одним із шляхів подолання вказаних недоліків є розробка інтелектуальної пошукової системи, яка інтегрує різні моделі когнітивної психології в єдину концептуальну модель представлення знань і комплексно використовує її під час інтерпретації.   1. Показано можливість формалізації концептуальної нечіткої гібридної моделі представлення знань на основі апарату нечітких множин. Це дозволяє автоматизувати інформаційні процеси, що моделюють людське мислення, а також врахувати індивідуальні знання про оточення і неявні знання, що містяться в текстових фрагментах. 2. Запропоновано інтерпретацію тексту розбити на етапи, які включають термінологічний, фразеологічний і концептуальний аналізи, що дозволяє автономно та незалежно формалізувати перебування тексту на різних етапах обробки 3. Етап концептуального аналізу побудований на основі моделі активності мереж, згідно з якою смисл тексту формується з урахуванням аналізованого фрагменту тексту, пам'яті і контексту. Це дозволило враховувати смислову зв'язність тексту і локалізацію знань у ньому під час побудови контексту. В основу моделі активності мережі покладена формалізована когнітивна модель логогену Мортона, адаптована для обробки тексту, на підставі НГМ. 4. Показано, що інтерпретацію тексту на основі нечіткої гібридної моделі представлення знань та моделі активності мереж можливо розглядати як формалізований механізм міркування на основі досвіду. 5. Приховані знання (неявні знання або смисл), здобуті шляхом висновку на основі досвіду, можуть бути використані як факти в нечітких продукційних системах. Це дозволяє, по-перше, інтегрувати на загальній математичній платформі два механізми висновку (висновок на основі прихованих і висновок на основі поверхових знань), а, по-друге, розширити область інтерпретації тексту, і, як наслідок, приймати рішення з більшим ступенем упевненості при використанні в автоматизованих системах категоризації, анотації, а також в ІПС. 6. Метод й моделі представлення та інтерпретації інформації реалізовані у вигляді бібліотеки процедур, яка дозволяє компонувати на її основі ядро АСІТ – автоматизовану систему категоризації та інтерпретації текстової інформації «Text-Term-Concept» – і налаштовувати на конкретний додаток шляхом формування інтенсіональної частини ГМ. Система використана для розробки інтелектуальної ІПС, яка пройшла дослідну експлуатацію на сайті Донецького національного університету. 7. АСІТ у вигляді інтелектуальної ІПС «ENewsAnalyzer» упроваджена на підприємстві ТОВ «УФС». За наслідками дослідної експлуатації системи «ENewsAnalyzer», до складу якої входить модуль отримання знань «SemanticExpert», модуль автоматизованого збору і передобробки фінансових даних «Intelligent*Mail* *Agent*», модуль інтерпретації текстової інформації та формування даних для продукційних правил «*TC*», отримана оцінка ефективності розробленого методу й моделей: обсяг актуальної (необхідної) для прийняття рішення інформації зростає від 20% до 65%. | |