

породження дискретних об'єктів: частиномовна розмітка (part-of-speech tagging), морфологічна сегментація, стемінг (stemming), виділення меж речення (sentence boundary disambiguation), розпізнавання іменованих сутностей, усунення невизначеності значень слів (word sense disambiguation), синтаксичний аналіз інформації (syntactic parsing), вирішення питання кореференцій (coreference resolution). До другого класу відносяться задачі, які вимагають розуміння тексту. Для таких задач легко запропонувати метрики, які не викликає сумнівів: створення мовних моделей, інформаційного пошуку, аналізу тональності, виділення відношень або фактів, відповідей на запитання. До третього класу належать задачі, які породжують новий текст на основі існуючого, та в яких метрики якості не завжди є очевидними. До таких задач відносяться задачі породження тексту, автоматичного реферування, машинного перекладу, діалогові моделі.

#### 39. АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ В ОБЛАСТИ ВЕРИФИКАЦИИ УСТРОЙСТВ НА БАЗЕ ПРОЦЕССОРОВ АРХИТЕКТУРЫ ARM

Корниенко В.Р., к.т.н. доц. Филиппенко И.В., ХНУРЕ, Харьков

В докладе рассмотрены основные методы тестирования и верификации микроконтроллерных систем, основанных на ядре ARM. Рассмотрены существующие решения в области отладки и проведения автоматического тестирования встраиваемых систем. Проведен анализ устройств для тестирования решений на базе ARM-FPGA SoC. Выявлены возможные недостатки существующих решений при тестировании граничных возможностей системы и их возможные варианты устранения. Предложена оптимальная архитектура проекта устройства для обеспечения возможности корректной верификации и дальнейшего сопровождения.

#### 40. РАСПОЗНАВАНИЕ СТРУКТУРЫ ТЕКОВЫХ ДОКУМЕНТОВ, ПРЕДСТАВЛЕННЫХ В ФОРМАТЕ PDF

к.т.н. доц. Чалая Л.Э., Гринев С.А., ХНУРЕ, Харьков

В докладе представлен метод, позволяющий распознавать текстовые и графические объекты в документах, представленных в формате PDF, с целью их дальнейшего сравнения или поиска по коллекции таких документов. В предлагаемом подходе распознавание структуры разделено на сегменты: распознавание авторов, заглавия, аннотации, собственно текста, расположенных в тексте графических объектов, колонтитулов. В однотипных текстовых документах (например, научных статьях) выделяется определенный порядок форматирования, некоторые подструктуры легко выделить при помощи шаблонов (с применением регулярных выражений). Для графических элементов определяется их тип: формула, рисунок, схема, график. На выходе программного модуля, реализующего предложенный метод, формируется XML документ, размеченный соответствующими тегами, что облегчает дальнейшую обработку и улучшает качество информационного поиска по коллекции документов.

#### 41. КЛАССИФИКАЦИЯ ГРАФИЧЕСКИХ ИЗОБРАЖЕНИЙ С ПРИМЕНЕНИЕМ МЕТОДА ОПОРНЫХ ВЕКТОРОВ

к.т.н. Волощук Е.Б., к.т.н. доц. Чалая Л.Э., Ширманов В.В., ХНУРЕ, Харьков

В докладе предложены метод и программное приложение для решения задачи классификации графических изображений. Приложение основано на использовании метода опорных векторов (МОВ) и многослойного перцептрона (МСП). Оба алгоритма показали точность более 80%, но все же машина опорных векторов была эффективнее на 6% в последнем сравнении. Тестирование показало, что машина опорных векторов обучается быстрее МСП для большинства обучающих выборок: при проведении классификации с извлечением SIFT параметров машина опорных векторов обучалась 1.39 сек, а МСП – 15.66 сек. При извлечении данных на основе радиального сканирования получены следующие результаты: время обучения МОВ – 0.26 секунды, МСП – 6 секунд (для выборки Swedish

leaf dataset). Следует также отметить, что в МОВ используется меньший набор параметров, которые нужно вручную настраивать для получения хороших значений точности классификации. Наилучшие результаты классификации графических изображений по всем обучающим выборкам получены для МОВ с линейным ядром.

#### 42. СЕРВИС ДЛЯ ВИЯВЛЕННЯ БОТІВ У СОЦІАЛЬНИХ МЕРЕЖАХ

Гриньова О.Є., Кисля А.Г., ХНУРЕ, Харків

З розвитком інформаційних технологій з'являється все більше способів маніпуляції людською свідомістю. Існує безліч інструментів і різних платформ, але саме соціальні мережі стали головним плацдармом для подібних дій. Соціальні боти – спеціальні програми, створені для імітації поведінки людей в соцмережах. Призначення цих програм може бути різним, але найчастіше вони використовуються інтернет-маркетологами та кіберзлочинцями. В доповіді розглядається можливість розпізнавання ботів з використанням методів машинного навчання. Існує чимало різних ознак, за якими можна розпізнати бота. Найчастіше боти не викладають велику кількість фотографій або новин на своїх сторінках, адже їх головна мета – докласти найменших зусиль для імітації роботи в соцмережі реального користувача. Боти часто приймають запити в друзі, але не можуть відповідати на повідомлення (або їхні відповіді обмежені певним набором фраз). Розроблений сервіс дозволяє виявляти ботів за результатами обробки масивів з повідомленнями та профілями ботів. В сервісі використовується алгоритм обчислень неявних ознак бота за допомогою нейронної мережі.

#### 43. ЗАБЕЗПЕЧЕННЯ БЕЗПЕКИ ДАНИХ В ХМАРНИХ СЕРВІСАХ

к.т.н. доц., Кобзев І.В., Петрова К.К., ХНУРЕ, Харків

Впровадження нових інформаційних технологій в економічну сферу дозволяє істотно покращувати схеми реалізації бізнес-процесів підприємств. Одним з актуальних напрямів такого покращання є використання хмарних сервісів. Для контролю стану хмарного середовища потрібно розробити ефективні моделі управління ризиками. Реалізація апаратної безпеки вимагає суворого контролю доступу до обчислювальних модулів. Однією з найважливіших умов стабільності роботи середовища є розмежування прав і забезпечення прозорості (непомітності застосування засобів контролю/захисту для зовнішніх підключень). Класичне представлення периметра мережі, що розділяє внутрішні і зовнішні сегменти, в хмарних рішеннях розмивається. Інфікування або несанкціонований доступ до однієї з віртуальних машин може спричинити ушкодження і інших елементів мережі. В доповіді запропоновано метод захисту від цієї загрози за допомогою міжмережного екрану, який використовує систему фільтрів, що розмежують внутрішні віртуальні обчислювальні машини від взаємного впливу.

#### 44. ВИЗНАЧЕННЯ ДЖЕРЕЛА АТАКИ В МЕРЕЖІ INTERNET ЗА IP-АДРЕСОЮ ПОРУШНИКА

к.н.держ.упр. доц. Онищенко Ю.М., к.т.н. доц. Гнусов Ю.В., ХНУВС, Харків

У доповіді розглянуто основні проблемні питання встановлення джерела атаки в мережі Інтернет за IP-адресою користувача. Під час роботи в мережі Інтернет користувач залишає на сайтах свою IP-адресу, дані про систему, браузер та іншу службу інформацію. Способи встановлення справжньої IP-адреси залежать від того, які способи приховування застосовував порушник. Для відстеження порушника, який використовує ланцюг різноманітних серверів, необхідно послідовно посилати запити власникам кожного сервера з метою встановлення наступної ланки в ланцюгу. Якщо в процесі відстеження порушника розслідування виходить за межі однієї держави, то офіційну взаємодію необхідно проводити із залученням міжнародних правоохоронних організацій. Це ускладнює процес і зменшує вірогідність того, що власник проху-сервера зможе оперативно надати інформацію про того, ким ця адреса була використана в момент атаки,

відповідно, ймовірність того, що порушника вдасться відстежити, є досить низькою. Навіть у разі використання ланцюжка з шести-семи серверів, розташованих на території різних держав, розслідування може зайняти не менше декількох місяців.

#### 45. ПРО ОЦІНЮВАННЯ НЕСТАЦІОНАРНИХ ПАРАМЕТРІВ, ЩО ОПИСУЮТЬСЯ МАРКІВСЬКОЮ МОДЕЛЛЮ

д.т.н. проф. Руденко О.Г., д.т.н. проф. Безсонов О.О., Романюк О.С., ХНЕУ, Харків

У доповіді розглядається задача оцінювання нестационарних параметрів, що представляють собою стаціонарні випадкові процеси, які можуть бути описані модифікованою марковською моделлю першого порядку. Для оцінювання параметрів використовується регуляризований алгоритм Качмажа. Застосування другого методу Ляпунова дозволило вперше отримати неасимптотичні та асимптотичні оцінки для даного алгоритму. Отримано співвідношення, що характеризують умови збіжності алгоритму в середньому та середньоквадратичному, з яких, як частний випадок, витікають оцінки для класичного алгоритму Качмажа для стаціонарного і нестационарного випадків. Проведений аналіз показав, що отримані результати більш точно відображають властивості алгоритма у порівнянні з існуючими оцінками.

#### 46. ГИБРИДНИЙ МЕТОД РАНЖИРОВАНИЯ РЕЗУЛЬТАТОВ ЗАПРОСОВ В ПОИСКОВЫХ СИСТЕМАХ

к.т.н. доц. Чалая Л.Э., Грачевский Д.С., ХНУРЭ, Харків

В докладе представлено web-приложение, основанное на гибридном методе ранжирования результатов запросов в поисковых системах. Метод заключается в ранжировании результатов последовательных запросов, отличающихся друг от друга заменой, добавлением или перестановкой слов. Результаты последовательно выполненных запросов маркируются (запросам присваиваются номера или другие метки по желанию пользователя) таким образом, чтобы пользователь имел возможность на одном экране просматривать историю ответов на свои запросы. Метод гибридного ранжирования целесообразно применять при исследовательском поиске, где важна релевантность результатов запросов. Показано, что наибольшую релевантность имеют те документы, которые при последовательных запросах получили большее количество меток.

#### 47. СИСТЕМА ПЕРЕКЛАДУ ТЕКСТУ З САМОНАВЧАЛЬНОЮ ЗДАТНІСТЮ

к.т.н. доц. Сумцов Д.В., Юношев М.О., Гмиря В.В., ХНУРЕ, Харків

На даний момент є багато різних систем перекладу тексту, як у вигляді програмного забезпечення, так і у вигляді веб-сервісів. Особливістю запропонованої авторами системи є самонавчальна здатність. Самонавчальна здатність забезпечена за допомогою статистичного методу. Статистичний метод включає сукупність прийомів, користуючись якими статистика досліджує предметну область. Метод статистики об'єднує в собі три групи методів: метод масових спостережень, метод угруповань, метод узагальнюючих показників. Запропонована система заснована на методі масових спостережень через його широке поширення та великий досвід використання. Розроблене програмне забезпечення системи надає користувачам можливість самостійно обирати найбільш прийнятний для них метод вивчення мови та має широкий потенціал для подальшого розвитку.

#### 48. ПРОГНОЗУВАННЯ ЗАПИТІВ КОРИСТУВАЧА ДО ІНФОРМАЦІЙНОЇ СИСТЕМИ

Гриньова О.С., Кузнецов О.В., ХНУРЕ, Харків

Дії користувачів всесвітньої павутини можна вдало прогнозувати з використанням нейронних мереж та сучасних операційних систем. В доповіді розглядається можливість такого прогнозування в разі використання останньої версії операційної системи android. Запропоновано процедуру прогнозування типу додатків, що є найбільш

актуальними у поточний час. Для компаній, які оперують великими об'ємами даних, можна стверджувати, що на очікування результату запиту завжди витрачається більше часу, ніж для створення цього запиту. Щодо самих запитів деякі критерії можна спрогнозувати (вибірка за поточний місяць, вибірка серед усіх даних доступних лише цьому користувачеві тощо), а час роботи майже завжди буде однаковий. Це дозволяє прогнозувати тип запиту за деякий час до того, як він буде складений, і зберегти результат його виконання. Якщо користувач має необхідність термінового отримання результату цього запиту, то його можна буде отримати миттєво з пам'яті системи. В доповіді наведено результати тестування запропонованої процедури прогнозування.

#### 49. СИСТЕМА АНАЛИЗА ТРАФИКА СЕТИ С ИСПОЛЬЗОВАНИЕМ ОБЪЕКТНО-ОРИЕНТИРОВАННОГО ПРОГРАММИРОВАНИЯ

к.т.н. доц. Затхей В.А., к.т.н., доц. Тесленко О.В., ХНЭУ, Харків

В докладе предлагается вариант модифицированной системы анализа трафика пользователей локальных сетей (TRAFFIC-M) на базе системы StarGazer и средств объектно-ориентированного программирования. Система TRAFFIC-M предназначена для учета трафика в локальных сетях и представляет собой модифицированную версию базовой системы StarGazer, дополненную созданными с использованием Ruby on Rails веб-приложениями. Система позволяет осуществлять: контроль и текущие корректировки трафика клиентов сети; авторизацию пользователя с последующим разрешением или запретом доступа в Internet; анализ трафика по заданным направлениям и правилам; оперативное определение израсходованных пользователем средств и автоматическое отключение в случае превышения лимита; автоматический пинг пользователей сети и вывод результатов на экран; оперативное предоставление пользователю информации о его трафике; формирование подробных отчетов о состоянии трафика пользователей; реализацию перечисленных услуг с использованием наглядного интерфейса пользователя

#### 50. ОПРЕДЕЛЕНИЕ УРОВНЯ РАЗРАБОТЧИКА ПРИ ПОМОЩИ ДЕРЕВА ПРИНЯТИЯ РЕШЕНИЙ

д.т.н. проф. Смеляков К.С., Иванов А.А., ХНУРЭ, Харків

Рассмотрена классификация специалиста определенного уровня в IT отрасли: trainee, junior, middle или senior используя заданную модель дерева принятия решений. При построении дерева принятия решений, был использован алгоритм ID3 (Iterative Dichotomiser 3). Этот алгоритм основан на том чтобы взять не использованные признаки и посчитать их энтропию относительно собранных данных, выбрать признак, для которого энтропия минимальна, а информационная выгода максимальна, и создать узел дерева, содержащий этот признак. В результате получаются избыточно детализированное дерево, и количество ошибок при дальнейшем использовании такого дерева растет. Для решения этой проблемы применяют усечение, рассматривая каждую вершину как кандидата на усечение. Поддеревом с корнем в этом узле удаляется из дерева, а в узел помещается метка с тем исходом, которых в этом узле большинство. В итоге проанализирован работу алгоритма, на статистически значимой выборке данных, с учетом мнения экспертов в предметной области, установлено, что предложенная усеченная модель является адекватной и позволяет эффективно выявлять уровень специалиста.

#### 51. ГНУЧКА ІНСТРУМЕНТАЛЬНО-ТЕХНОЛОГІЧНА ПЛАТФОРМА ІНТЕГРАЦІЇ КОМПОНЕНТІВ ЗАСТОСУВАНЬ РІВНЯ ПІДПРИЄМСТВА

к.т.н. доц. Борисенко В.П., ХНУРЕ, Харків

В роботі запропонована гнучка інструментально-технологічна платформа (ГП) – як єдиний набір інструментальних засобів, об'єднаних в загальну архітектуру на основі базової технології веб-сервісів. ГП забезпечує зниження загальної вартості володіння корпоративними інформаційними системами (Total Cost of Ownership).