

МЕТОДИ ЗА СЪБИРАНЕ, ОРГАНИЗАЦИЯ, ДОСТЪП, ОБРАБОТКА И АНАЛИЗ НА ГОЛЕМИ ОБЕМИ ОТ СТРУКТУРИРАНИ И НЕСТРУКТУРИРАНИ ДАННИ

ПРОЕКТ 2018-ФЕЕА-01

Тема на проекта:
МЕТОДИ ЗА СЪБИРАНЕ, ОРГАНИЗАЦИЯ, ДОСТЪП, ОБРАБОТКА И АНАЛИЗ НА ГОЛЕМИ ОБЕМИ ОТ СТРУКТУРИРАНИ И НЕСТРУКТУРИРАНИ ДАННИ

Ръководител:
доц. д-р инж. Ирена Маринова Вълвова

Работен колектив:
доказани учени от катедра КСТ - 18; докторанти и постдокторанти - 13 студенти - 10

Адрес: 7017 Русе, ул. "Студентска" 8, Русенски университет "Ангел Кънчев"
Тел.: 082 - 888 685
E-mail: ivalova@ecs.uni-ruse.bg

Цел на проекта:
Да се изследват съществуващите и ако е необходимо да се предложат нови методи и средства за извличане, анализ и обработка на големи обеми неструктурирани и структурирани данни, в това число текст, документи, изображения, данни, събирани от интелигентни сензори в реално време и други. Необходимо е също да се предложат и изследват различни методи и алгоритми за извличане на специфични характеристики и метаданни. Целта е получените данни да се обработват и анализират автоматизирано, в резултат на което да се изготвят краткосрочни и дългосрочни прогнози и анализи в съответните сфери на дейност.

Основни задачи:

- Проучване, анализиране и използване на подходящи алгоритми за автоматично извличане на характеристики (метаданни) от тях;
- Анализ на характеристиките на нерелационните модели на данните;
- Проучване и анализ на мерки за сходство и алгоритми за търсене;
- Анализиране и използване на подходящи алгоритми за събиране, анализ и структурирано съхранение на големи обеми данни, потребителски профили и данни, извлечени от социални мрежи;
- Създаване на модели за организация на съхранението на тези данни в свързани големи разпределени БД;
- Проучване на съществуващите технологични решения за обучение на хора със специални образователни потребности (СОП);
- Провеждане на експериментални изследвания и анализи.

Основни резултати:

- Изследвани са различни характеристики за описване на обекти и мерки за сходство на неструктурирани данни и алгоритми за автоматично извличане на метаданни;
- Анализиран са характеристиките на нерелационните модели на данните;
- В рамките на проекта са направени редица изследвания и са оценени предимствата и недостатъците на различните решения, както и са предложени иновативни решения за приобщаване образование на хора със специални образователни потребности.

Публикации:

- 1 публикация в списание с импакт-фактор/импакт-ранг
- 15 публикации в Scopus/Web of Science
- 14 доклада в СНС, 1 - в НК на РУ, 5 - в други научни конференции, 3 от които международни

Други:

АНОТАЦИЯ

В резултат на човешката дейност в редица области от живота се генерират голямо количество несвързани документи, които обикновено се съхраняват само на хартия. Хартиеният носител не позволява да се извършва бързо и лесно търсене, а от там и статистически анализ на натрупаните през годините данни. Дори и в случая на електронни документи, липсата на взаимовръзки между тях силно ограничава гледната точка на експертите и не позволява да се провеждат достатъчно задълбочени и широкообхватни анализи и прогнози. В резултат крайните изводи и заключения на анализаторите често са непълни или недостатъчно прецизни, което води след себе си и до неточни и/или неефективни стратегии за управление на процесите от съответните области.

Съвременните информационни и комуникационни технологии биха могли да имат съществен принос в решаването на този проблем. Преминването от хартиени към електронни документи ще позволи тяхното последващо автоматизирано обработване. Ако върху тях се приложат различни методи и алгоритми за семантичен анализ, то несвързаните на пръв поглед документи биха представили една много по-пълна и богата информация за процесите, за които се отнасят. Последващият семантичен и статистически анализ на тази информация ще позволи автоматизирано да се правят изводи и да се откриват тенденции, а от там и да се планират и оптимизират съответните процеси.

Същевременно в последните години се наблюдава рязко увеличаване на обема на съхраняваните колекции от изображения. Ежедневно се генерират стотици и хиляди гига байти от изображения за най-различни цели и приложения. Тази информация трудно би могла да се използва пълно, ако не е организирана по подходящ начин, така че да дава възможност за ефикасно разглеждане, търсене и извличане на различни заявки за достъп.

В наши дни все по-актуални стават различните варианти на интелигентни решения за дома, офиса или другите сфери от ежедневието ни. При разработването и използването на такива решения също се изисква събиране на огромно количество данни от различни сензори, съхраняването им, анализ и обработка, с цел прогнозиране и интелигентно управление на различни обекти.

Развитието на съвременните информационни и комуникационни технологии предоставя все по-голям потенциал и за разработване на решения, представящи помощ и подкрепа за деца и хора със сензорни и двигателни увреждания, деца с нарушения на аутистичния спектър, със специфични обучителни трудности и др. Тези технологични решения могат да спомогнат за осигуряването на равен достъп за хора в неравностойно положение до масовото образование, както и за подобряването на качеството им на живот в ежедневието. При деца със специални образователни потребности, компютърните технологии могат да предоставят възможности за съдействие при преодоляване на барьерите при обучаването, а при хора с увреждания е налице потенциал за създаване на решения за по-добро ориентиране и придвижване в околната среда, за известяване чрез употребата на различни сензорни стимули и др.

PROJECT 2018-FEEA-01

Project title:
METHODS OF COLLECTION, ORGANIZATION, ACCESS, PROCESSING AND ANALYSIS OF LARGE VOLUMES OF STRUCTURED AND NON-TRUSTED DATA

Project director:
Assoc. Prof. Irena Valova, PhD

Project team:
researchers/professors from CST department - 18; PhD and post PhD - 13 students - 10

Address: University of Ruse, 8 Studentska str., 7017 Ruse, Bulgaria
Phone: +359 82 - 888 685
E-mail: ivalova@ecs.uni-ruse.bg

Project objective:
Review of existing and proposing new (if necessary) methods and tools for extracting, analyzing and processing large volumes of unstructured and structured data, including text, documents, images, data collected by smart sensors in real time, and others. Development and experimental analysis of new methods and algorithms for feature and metadata extraction. All data should be processed automatically and results could be used for preparing short term analyses and long term forecasts in various subject domains.

Main activities:

- Research, analysis and utilization of algorithms for automatic feature and metadata extraction;
- Development of methods for data organization and retrieval;
- Analysis of non-relational data models and their use in NoSQL DBMS implementations;
- Research and analysis of similarity measures and search algorithms;
- Analysis and implementation of algorithms for collection, analysis, and structural storage of large volumes of data fetched from sensors or other sources;
- Development of new models for data organization and storage in over-sized distributed databases;
- Research, analysis and implementation of algorithms for automatic user profiling and preference analysis based on data retrieved from social networks;
- Research of existing software and hardware technology solutions for training people with special educational needs (SEN);
- Conducting experimental studies and analyses.

Main outcomes:

- Different features to describe objects have been explored;
- A study of existing similarity measures of unstructured data;
- Analysis of various metadata retrieval algorithms;
- Analysis of non-relational data models and their use in NoSQL DBMSs.

Publications:

- 1 paper sent to the scientific journal with Impact factor
- 15 papers published in Scopus/Web of Science
- 14 papers in Students Scientific Conference, 1 - in the Scientific Conference in RU, 5 - in other Scientific Conferences, 3 of them are International

Others:

ИЗСЛЕДВАНЕ И АНАЛИЗ НА РЕЗУЛТАТИТЕ ОТ ОЦЕНКАТА НА ПОДОБИЕТО МЕЖДУ ИЗОБРАЖЕНИЯ ЧРЕЗ РАЗЛИЧНИ МЕРКИ ЗА ПОДОБИЕ

Query Image	Search Image	Jaccard	Euclidean Distance	Cityblock Distance	Chi-square Dissimilarity	Kullback-Libler Divergence	Jeffrey Divergence	My Array Differences
		100.0000	(1)	100.0000	(1)	100.0000	(1)	100.0000
		74.1211	(6)	89.0470	(4)	50.8475	(3)	99.5626
		69.4336	(54)	84.1797	(35)	30.7825	(40)	99.5385
		47.2656	(150)	75.0667	(129)	78.0120	(144)	98.8977
		44.2383	(31)	79.4054	(63)	66.9959	(62)	98.8865
		43.9453	(172)	71.6967	(177)	74.5749	(178)	98.7637
		21.4844	(215)	54.2814	(216)	63.6046	(216)	98.5257
		21.0938	(213)	54.3728	(215)	58.7269	(215)	98.0886
		17.5781	(214)	54.1341	(214)	76.1851	(214)	98.2184

ИНФОРМАЦИОННА СИСТЕМА ЗА ПУБЛИКАЦИОННАТА ДЕЙНОСТ В РУСЕНСКИ УНИВЕРСИТЕТ

ИНОВАТИВЕН МЕТОД ЗА ИЗУЧАВАНЕ НА КОМПОНЕНТИТЕ НА ДЪННАТА ПЛАТКА С ДОБАВЕНА РЕАЛНОСТ

3D модели на компютърни компоненти

Визуализиране на модел на процесор върху маркер

Прилагане на скалиране върху модела

Създаване и изследване на технологични решения за приобщаване образование на деца със специални образователни потребности

Интерактивни обучителни инструменти

Взаимовръзки в процеса на създаване на обобщен профил на обучаваното дете със СОП или СОТ в веб базиратата среда

Диаграма на дейности на преподавател

Диаграма на дейности на обучаван