

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. Иrena Marinova Вълова

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 извлечане, анализ и обработка на големи обеми неструктурирани и структурирани данни, в това число текст, документи, изображения, данни, събиранни от интелигентни сензори в реално време и други. Необходимо е също да се предложат и изследват различни методи и алгоритми за извлечане 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 извлечане 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 свръх 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 описание на обекти и мерки за сходство на неструктурирани данни и алгоритми за автоматично извлечане на метаданни ;
- Анализирани 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 приобщаващо 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. Преминаването от хартиени към електронни документи ще позволи тяхното последващо автоматизирано обработване. А 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 генерираят стотици и хиляди гигабайти 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 изпълнение на различни заявки за достъп.

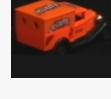
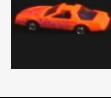
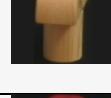
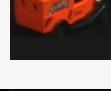
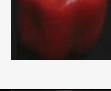
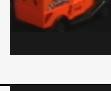
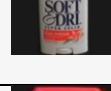
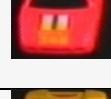
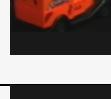
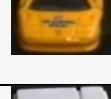
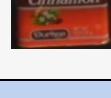
В наши дни все по-актуални стават различните варианти на интелигентни решения за дома, офиса или другите сфери от ежедневието ни. При разработването и използването на такива решения също се изисква събиране на огромно количество данни от различни сензори, съхраняването им, анализ и обработка, с цел прогнозиране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

Развитието на съвременните информационни и комуникационни технологии предоставя все по-голям потенциал и за разработване на решения, представляващи помощ и подкрепа за деца и хора със сензорни и двигателни увреждания, деца с нарушения от 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 бариерите при общуването, 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

<p>Project title: METHODS OF COLLECTION, ORGANIZATION, ACCESS, PROCESSING AND ANALYSIS OF LARGE VOLUMES OF STRUCTURED AND NON-TRUSTED DATA</p>
<p>Project director: Assoc. Prof. Irena Valova, PhD</p>
<p>Project team: researchers/professors from CST department - 18; PhD and post PhD - 13 students - 10</p>
<p>Address: University of Ruse, 8 Studentska str., 7017 Ruse, Bulgaria Phone: +359 82 - 888 685 E-mail: ivalova@ecs.uni-ruse.bg</p>
<p>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.</p>
<p>Main activities:</p> <ul style="list-style-type: none">• 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.
<p>Main outcomes:</p> <ul style="list-style-type: none">• 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.
<p>Publications:</p> <ul style="list-style-type: none">• 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
<p>Others:</p> <ul style="list-style-type: none">•

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-liebler Divergence	Jeffrey Divergence	My Array Differences
		(1) 100.0000	(1) 100.0000	(1) 100.0000	(1) 100.0000	(1) 100.0000	(1) 100.0000	(1) 100.0000
		(2) 74.1211	(6) 75.3952	(2) 89.0470	(4) 91.8441	(150) 50.8475	(3) 92.2930	(4) 99.5626
		(3) 69.4336	(54) 68.5930	(27) 84.1797	(35) 86.7444	(193) 30.7825	(40) 87.2650	(5) 99.5385
		(96) 47.2656	(150) 59.5795	(129) 75.0667	(139) 78.0120	(126) 58.8525	(144) 78.5414	(81) 98.8977
		(125) 44.2383	(31) 70.6438	(74) 79.4054	(63) 84.3367	(90) 68.9959	(62) 85.4059	(84) 98.8865
		(126) 43.9453	(172) 56.9767	(162) 71.6967	(177) 74.5749	(148) 51.8022	(178) 75.1894	(113) 98.7637
		(214) 21.4844	(215) 43.4418	(215) 54.2814	(216) 58.5133	(108) 63.6046	(216) 59.5224	(143) 98.5257
		(215) 21.0938	(213) 43.9430	(214) 54.3728	(215) 58.7269	(128) 58.6051	(215) 59.8004	(200) 98.0886
		(216) 17.5781	(214) 43.5530	(216) 54.1341	(214) 59.3517	(78) 73.1851	(214) 60.6498	(186) 98.2184

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 приобщаващо образование на деца със специални образователни потребности

The diagram illustrates the creation and development of technological solutions for inclusive education of children with special educational needs. It shows various stakeholders contributing to a central 'Обобщен профил на обучаемия' (General profile of the learner):

- ЛОГОПЕД** (Speech Therapist) contributes to the **ПРОФИЛ НА СПЕЦИАЛИСТА** (Profile of the specialist).
- ОБУЧАЕМ** (Student) contributes to the **ОБУЧИТЕЛЕН ПРОФИЛ** (Teacher profile).
- РЕСУРСЕН ПЕДАГОГ** (Resource Teacher) contributes to the **ПСИХОЛОГИЧЕСКИ ПРОФИЛ** (Psychological profile).
- УЧИТЕЛ** (Teacher) contributes to the **ОБОБЩЕН ПРОФИЛ НА ОБУЧАЕМИЯ** (General profile of the learner).
- ПСИХОЛОГ** (Psychologist) contributes to the **ПСИХОЛОГИЧЕСКИ ПРОФИЛ** (Psychological profile).
- РОДИТЕЛ** (Parent) contributes to the **ОБОБЩЕН ПРОФИЛ НА ОБУЧАЕМИЯ** (General profile of the learner).

Взаимовръзки в процеса на създаване на обобщен профил на обучаемото дете със СОП или СОТ в уеб базираната среда

Интерактивни обучителни инструменти

The screenshot shows an interactive educational tool for arithmetic operations. The interface includes:

- Logical operations:** Operation "AND" (Операция "И"), Operation "OR" (Операция "ИЛИ"), Operation "NOT" (Операция "Ф").
- Arithmetical operations:** Addition (Събиране), Subtraction (Изваждане), Multiplication (Умножение), Division (Деление).
- Binary System:** Choose a numeral system: binary (00)₁₆, hexadecimal (00)₁₆.
- Input Fields:** Enter remainder R: 0; Enter C(R)mod3: 0; Enter C(Q)*C(B): 1; Enter C(Q)*C(B): 1; Enter C(Q)*C(B)mod3: 1; Enter C(Q)*C(B)mod3: 1; Enter C(Q)*C(B)mod3: 1; Enter C(Q)*C(B)mod3: 1.
- Output Fields:** A=(1)₂; B=(1)₂; Q=(1)₂; C(R)=(0)₂; C(R)mod3=(0); C(Q)*C(B)=(1), C(Q)*C(B)=(1), C(C(Q)*C(B)mod3 + C(R))=(1), C(C(Q)*C(B)mod3 + C(R))=(1).
- Buttons:** choose, Next.

The screenshot shows another version of the interactive educational tool for arithmetic operations. The interface includes:

- Logical operations:** Operation "AND" (Операция "И") with formulae $C(A,B) = [C(A) + C(B) - C(A \cdot B)] \text{ mod} 3$ and $C_2(A,B) = [C(A) + C(B) - C(A \oplus B)] \text{ mod} 3$.
- Arithmetical operations:** Addition (Събиране), Subtraction (Изваждане), Multiplication (Умножение), Division (Деление).
- Binary System:** Choose a numeral system: binary (A=03)₁₆, hexadecimal (B=(03)₁₆).
- Input Fields:** Enter OperandsA: 1 1; Enter OperandsB: 1 1; Enter A,B: 1 1; Remainder in mod3 0 of A,B: 0; Remainder in mod3 0 of operand A: 0; Remainder in mod3 0 of operand B: 0.
- Output Fields:** A-B=(03)₁₆; C(A,B)=(00)₁₆; C(A)=(00)₁₆.
- Buttons:** choose, Next.

Диаграма на дейности на преподавател

```

graph TD
    A((Избран е модул "Ръководство")) --> B((Избран е модул "Задача"))
    B --> C[Решаване на задача]
    C --> D[Съобщение за грешка]
    D --> E[Въвеждане на пример за решаване]
    E --> F[Извриване/промяна на пример]
    F --> G[Избор на тип грешка]
    G --> H[Избор на пораждащ полином от преизчислен списък за конкретния пример]
    H --> I[Извеждане на: - информационни разряди - контролни разряди - кодово разстояние - тип на грешката - k, d, l0, m, n - G(x), xk.G(x), P(x)]
    I --> J[Извеждане на новообразуващият се пораждащ полином P(x) = (1+x).P1(x)]
    J --> K[Избор на модул "Ръководство"]
    K --> L[Предвид на ръководство]
    L --> M[Решаване на задача]
    M --> N[Съобщение за грешка]
    N --> O[Въвеждане на пример за решаване]
    O --> P[Извриване/промяна на пример]
    P --> Q[Избор на тип грешка]
    Q --> R[Избор на пораждащ полином от преизчислен списък за конкретния пример]
    R --> S[Извеждане на: - информационни разряди - контролни разряди - кодово разстояние - тип на грешката - k, d, l0, m, n - G(x), xk.G(x)]
    S --> T[Извеждане на новообразуващият се пораждащ полином P(x) = (1+x).P1(x)]
    T --> U[Избор на модул "Задача"]
    U --> V[Предвид на задача]
    V --> W[Решаване на задача]
    W --> X[Съобщение за грешка]
    X --> Y[Въвеждане на пример за решаване]
    Y --> Z[Извриване/промяна на пример]
    Z --> AA[Избор на тип грешка]
    AA --> BB[Избор на пораждащ полином от преизчислен списък за конкретния пример]
    BB --> CC[Извеждане на: - информационни разряди - контролни разряди - кодово разстояние - тип на грешката - k, d, l0, m, n - G(x), xk.G(x)]
    CC --> DD[Извеждане на новообразуващият се пораждащ полином P(x) = (1+x).P1(x)]
    DD --> EE[Избор на модул "Ръководство"]
    EE --> FF[Предвид на ръководство]
    FF --> GG[Решаване на задача]
    GG --> HH[Съобщение за грешка]
    HH --> II[Въвеждане на пример за решаване]
    II --> JJ[Извриване/промяна на пример]
    JJ --> KK[Избор на тип грешка]
    KK --> LL[Избор на пораждащ полином от преизчислен списък за конкретния пример]
    LL --> MM[Извеждане на: - информационни разряди - контролни разряди - кодово разстояние - тип на грешката - k, d, l0, m, n - G(x), xk.G(x)]
    MM --> NN[Извеждане на новообразуващият се пораждащ полином P(x) = (1+x).P1(x)]
    NN --> OO[Избор на модул "Задача"]
    OO --> PP[Предвид на задача]
    PP --> QQ[Решаване на задача]
    QQ --> RR[Съобщение за грешка]
    RR --> SS[Въвеждане на пример за решаване]
    SS --> TT[Извриване/промяна на пример]
    TT --> AA
  
```

Диаграма на дейности на обучаван

```

graph TD
    A((Избран е модул "Теория")) --> B[Преглед на теория]
    B --> C[Решаване на задачи]
    C --> D[Извеждане резултат от умножение на xk.G(x)]
    D --> E[ЦИКЛ Извеждане резултат от умножение на xk.G(x) > степента на R(x) || R(xk.G(x)) != 0]
    E --> F[Извеждане на член в резултата от делението xk.G(x) / xk]
    F --> G[Извеждане на резултата от S=(xk.G(x) / xk) * P(x)]
    G --> H[Извеждане на резултата от S ⊕ R(xk.G(x))]
    H --> I[Извеждане на: - R(x) - F(x) = xk.G(x) + R(x) - xk.G(x) / xk]
    I --> J[Отчитане на грешка]
    J --> K[Въвеждане резултат от умножение на xk.G(x)]
    K --> L[Извеждане резултат от умножение на xk.G(x)]
    L --> M[Извеждане резултат от умножение на xk.G(x)]
    M --> N[Извеждане резултат от умножение на xk.G(x)]
    N --> O[Извеждане резултат от умножение на xk.G(x)]
    O --> P[Извеждане резултат от умножение на xk.G(x)]
    P --> Q[Извеждане резултат от умножение на xk.G(x)]
    Q --> R[Извеждане резултат от умножение на xk.G(x)]
    R --> S[Извеждане резултат от умножение на xk.G(x)]
    S --> T[Извеждане резултат от умножение на xk.G(x)]
    T --> U[Извеждане резултат от умножение на xk.G(x)]
    U --> V[Извеждане резултат от умножение на xk.G(x)]
    V --> W[Извеждане резултат от умножение на xk.G(x)]
    W --> X[Извеждане резултат от умножение на xk.G(x)]
    X --> Y[Извеждане резултат от умножение на xk.G(x)]
    Y --> Z[Извеждане резултат от умножение на xk.G(x)]
    Z --> AA[Извеждане резултат от умножение на xk.G(x)]
    AA --> BB[Извеждане резултат от умножение на xk.G(x)]
    BB --> CC[Извеждане резултат от умножение на xk.G(x)]
    CC --> DD[Извеждане резултат от умножение на xk.G(x)]
    DD --> EE[Извеждане резултат от умножение на xk.G(x)]
    EE --> FF[Извеждане резултат от умножение на xk.G(x)]
    FF --> GG[Извеждане резултат от умножение на xk.G(x)]
    GG --> HH[Извеждане резултат от умножение на xk.G(x)]
    HH --> II[Извеждане резултат от умножение на xk.G(x)]
    II --> JJ[Извеждане резултат от умножение на xk.G(x)]
    JJ --> KK[Извеждане резултат от умножение на xk.G(x)]
    KK --> LL[Извеждане резултат от умножение на xk.G(x)]
    LL --> MM[Извеждане резултат от умножение на xk.G(x)]
    MM --> NN[Извеждане резултат от умножение на xk.G(x)]
    NN --> OO[Извеждане резултат от умножение на xk.G(x)]
    OO --> PP[Извеждане резултат от умножение на xk.G(x)]
    PP --> QQ[Извеждане резултат от умножение на xk.G(x)]
    QQ --> RR[Извеждане резултат от умножение на xk.G(x)]
    RR --> SS[Извеждане резултат от умножение на xk.G(x)]
    SS --> TT[Извеждане резултат от умножение на xk.G(x)]
    TT --> AA
  
```