



P. Р. Лавценко, О. О. Водка

Національний технічний університет "Харківський політехнічний інститут", м. Харків, Україна

РОЗРОБЛЕННЯ АЛГОРИТМІВ ВІЗУАЛІЗАЦІЇ ВЛАСТИВОСТЕЙ ГЕТЕРОГЕННИХ ТА КОМПОЗИЦІЙНИХ МАТЕРІАЛІВ

Сьогодні у людській діяльності використовують дуже велику кількість різних матеріалів. Ці матеріали відрізняються між собою механічними властивостями, кольором, хімічним складом. Широке розмаїття матеріалів становить завдання про збереження інформації про них та про її візуалізацію. Створено базу даних матеріалів та веб-застосунок, який дає можливість отримати доступ до інформації про різні види металів та їхні сплави, а також підібрати найближчі аналоги за допомогою функції інтелектуального пошуку. Для створення бази даних використано систему керування базами даних MySQL. Наповнення цієї бази даних відбувалось способом розроблення спеціальної програми-парсера. Ця програма переглядала сторінки з мережі Інтернет та виділяла інформацію про властивості матеріалів і зберігала їх до бази даних. Для візуалізації даних використано алгоритми зниження розмірності. Ці алгоритми спрямовані на відшукування такої проекції високорозмірних даних на низькорозмірний простір. Під час використання цієї проекції зберігаються всі внутрішні взаємозв'язки між даними. Це дало змогу побудувати відображення великорозмірних даних на площину та візуалізувати їх. Для цього використано алгоритми Isomap, MDS, t-SNE. В такий спосіб розроблено базу даних матеріалів. Доступ до бази даних відбувається за допомогою веб-застосунку. Для зібраних даних побудовано візуалізацію. За результатами візуалізації виявлено кластери матеріалів, які відповідають вже відомим класам матеріалів. Це підтверджує коректність побудови моделей візуалізації та зниження розмірності, а також правильність зібраної інформації.

Ключові слова: веб-застосунок; метали; алгоритми зниження розмірності; багатомірні набори даних; інтелектуальний пошук; Isomap; MDS; t-SNE.

Вступ. Сьогодні в техніці використовують незліченну кількість різних матеріалів. Їхні фізичні та хімічні властивості варіюються у широкому спектрі. Однак найбільший, з-поміж інших чинників, вплив на всі властивості матеріалів має їхній хімічний склад. Саме він є визначальним для багатьох металів.

Метали і їхні сплави часто використовують для виготовлення інструментів, устаткування, різних машин, будівельних конструкцій, застосовують у промисловості і тощо. Отже, широке розмаїття металів не лише розкриває великі можливості для їх застосування.

До складу більшості сплавів може входити багато різних металевих та неметалевих елементів. Тому у складі сплавів можуть бути практично всі елементи з періодичної системи Менделєєва. Це становить завдання класифікації сплавів залежно від їх хімічного складу, властивостей і способу застосування у виробництві.

Подібні класифікації можуть займати дуже великі обсяги інформації і не мати якої-небудь структури, тому хорошим рішенням у цьому випадку може бути організація інформації в реляційну базу даних, наприклад, за допомогою СУБД MySQL (Diubua & Sher, 2005).

Хоча використання реляційної бази даних вирішує питання зручного зберігання наявної інформації, це не вирішує питання її зручного перегляду і швидкого дос-

тупу для усіх бажаючих. Веб-застосунок надає можливість його використання з усіх пристроїв, які мають вихід в Інтернет і позбуває від необхідності установки додаткового програмного забезпечення (Grinberg, 2018; Dwyer, Aggarwal & Stouffer, 2017; Maia, 2015; Aggarwal & Gupta, 2014).

Для візуалізації даних одним із головних завдань є зниження розмірності даних. Зниження розмірності потребує доступні способи візуалізації, які ефективно працюють у двох- чи трьохвимірному просторі.

Алгоритми зменшення розмірності базуються на думці, що розмірність багатьох наборів даних є лише штучно високою. Високорозмірні набори даних можуть бути дуже складними для візуалізації. Хоча дані за двома або трьома параметрами можна побудувати так, щоб показати властиву структуру даних, еквівалентні великогабаритні ділянки набагато менш інтуїтивно зрозумілі. Щоб допомогти візуалізувати структуру набору даних, розмір має бути певним чином зменшений.

Найпростішим способом досягнення зменшення розмірності є випадкове проектування даних. Хоча це дає змогу певним чином візуалізувати структури даних, випадковість вибору залишає бажати кращого. Під час використання випадкової проекції цілком імовірно, що структура в даних буде втрачена. Для вирішення цієї

Інформація про авторів:

Лавценко Руслан Ровшаногли, студент, кафедра динаміки та міцності машин. Email: vincitoremonster1402@gmail.com

Водка Олексій Олександрович, канд. техн. наук, доцент, кафедра динаміки та міцності машин. Email: oleksii.vodka@gmail.com

Цитування за ДСТУ: Лавценко Р. Р., Водка О. О. Розроблення алгоритмів візуалізації властивостей гетерогенних та композиційних матеріалів. Науковий вісник НЛТУ України. 2018, т. 28, № 11. С. 125–129.

Citation APA: Lavshchenko, R. R., & Vodka, O. O. (2018). Development of algorithms for visualizing the properties of heterogeneous and composite materials. *Scientific Bulletin of UNFU*, 28(11), 125–129. <https://doi.org/10.15421/40281122>

проблеми було розроблено алгоритми зменшення розмірності, такі як: аналіз основних компонентів (PCA, Jolliffe, 2011), багатомірного масштабування (MDS, Wickelmaier, 2003), стохастичного включення сусідів (t-SNE, SNE, Maaten, Van Der & Hinton, 2008; Hinton & Roweis, 2003).

Постановка завдання. У роботі необхідно створити базу даних матеріалів, що використовують у машинобудуванні. Для створення бази даних використати доступні відомості в мережі Інтернет. На основі отриманих даних побудувати візуалізацію матеріалів.

Створення бази даних матеріалів. На сьогодні відомо достатню кількість ресурсів, які надають інформацію про матеріали, тому збір інформації проводять на їхній основі.

Для цього завдання добре підходить синтаксичний аналізатор сторінок – парсер. У цій роботі використовували власний парсер, який збирав необхідні відомості про матеріали. Перший етап його роботи – це збір усіх сторінок із матеріалами, на яких надалі робитимемо синтаксичний аналіз. Для цього запускаємо порожній пошук і проходимо двома циклами по всіх посиланнях на кожній доступній сторінці пошуку (рис. 1,а). Внаслідок чого отримуємо масив із посиланнями на всі доступні матеріали. Тепер, маючи посилання на сторінки всіх матеріалів (див. рис. 1,б), можемо використати функцію синтаксичного аналізу для кожного з них.

Страницы: 1 2 3 4 5 6 7 8 9 10

1. [Сталь для отливок 10X18H3ГЭД2Л](#)
ль для отливок 10X18H3ГЭД2Л Марка: 10X18H3ГЭД2Л Класс: Сталь для отливок с особыми свойствами Использование в промышленности: для кавитационно-стойких деталей рабочей части гидротурбин, работающих при ...
2. [Сталь для отливок 10X18H11БЛ](#)
ль для отливок 10X18H11БЛ Марка: 10X18H11БЛ Класс: Сталь для отливок с особыми свойствами Использование в промышленности: для деталей, работающих при малых нагрузках; сталь жаропрочная до 800 град. С; ...
3. [Сталь для отливок 10X17H10Г4МБЛ](#)
ль для отливок 10X17H10Г4МБЛ Марка: 10X17H10Г4МБЛ Класс: Сталь для отливок с особыми свойствами Использование в промышленности: для коррозионно-стойких и жаропрочных деталей (литые детали) ...
4. [Сталь для отливок 10X14НДЛ](#)
ль для отливок 10X14НДЛ Марка: 10X14НДЛ Класс: Сталь для отливок с особыми свойствами Использование в промышленности: для деталей, коррозионно-стойких в морской воде и атмосферных условиях(гребные винты и ...
5. [Сталь для отливок 10X12НДЛ](#)
ль для отливок 10X12НДЛ Марка: 10X12НДЛ Класс: Сталь для отливок с особыми свойствами Использование в промышленности: для деталей, коррозионно-стойких и эрозионно-стойких в условиях проточной воды (детали ...
6. [Сталь для отливок 09X17H3СЛ](#)
ль для отливок 09X17H3СЛ Марка: 09X17H3СЛ Класс: Сталь для отливок с особыми свойствами Использование в промышленности: для деталей повышенной прочности, работающих в средах средней агрессивности (азотная ...
7. [Сталь для отливок 09X16H4БЛ](#)
ль для отливок 09X16H4БЛ Марка: 09X16H4БЛ Класс: Сталь для отливок с особыми свойствами Использование в промышленности: для коррозионно-стойких деталей повышенной прочности, устойчивых против окисления в ...

Механические свойства стали 08					
ГОСТ	Состояние поставки	σ_B (МПа)	δ_5 (%)	ψ (%)	НВ, не более
1050-88 10702-78	Сталь горячекатаная кованая, калиброванная и серебрянка 2-й категории после нормализации. Сталь калиброванная и калиброванная со специальной отделкой после отжига или отпуска.	325 315-410	33	60	---
1577-93	Сталь иагартванная калиброванная и калиброванная со специальной отделкой без термообработки.	370 310 320	---	55	131 179
10234-77 4041-71	Листы нормализованные и горячекатаные. Полосы нормализованные или горячекатаные. Лента отожженная плоская	Не более 450	20	33	---
16523-70	Лист термообработанный 1-2-й категории. Лист горячекатанный	275-410 275-390	32	---	109

Физические свойства стали 08						
T (град)	$E \cdot 10^{-5}$ (МПа)	$\alpha \cdot 10^{-6}$ (1/град)	λ (Вт/(м·град))	ρ (кг/м ³)	C (Дж/(кг·град))	R 10^9 (Ом·м)
20	2.03			7871		
100	2.07	12.5	60	7846	482	178
200	1.82	13.4	56	7814	498	252
300	1.53	14	51	7781	514	341
400	1.41	14.5	47	7745	533	448
500		14.9	41	7708	555	575
600		15.1	37	7668	584	725
700		15.3	34	7628	626	898
800		14.7	30	7598	695	1073
900		12.7	27	7602	703	1124
1000		13.8			695	

Рис. 1. Фрагмент сторінки ресурсу: а) з посиланнями на матеріали; б) властивості матеріалу

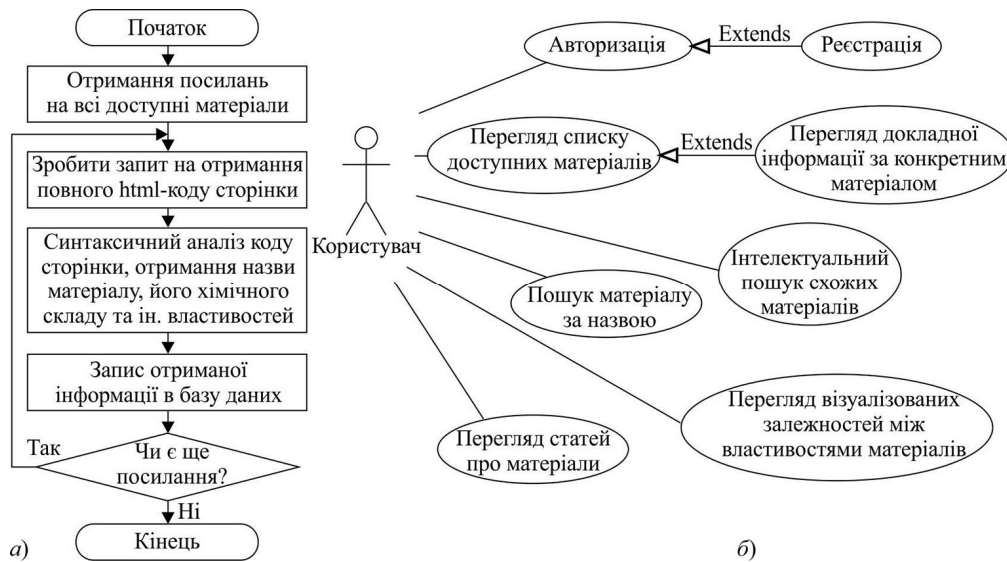


Рис. 2. Принципи роботи алгоритму: а) блок-схема алгоритму роботи парсера; б) діаграма прецедентів для користувача веб-застосунку

Аналізуючи структури сторінки, видно, що назва матеріалу поміщена в html-тег "h2". Шукаємо першу інформаційну таблицю і беремо звідти марку, клас і спосіб використання матеріалу. Наступним кроком знаходимо таблицю з хімічним складом матеріалу, потім проходимо в циклі по усіх комірках першої та другої колонки, збираючи в масив назви хімічних елементів та їхню концентрацію відповідно. Також відшукуємо таб-

лицю з різними властивостями і беремо з неї температуру критичних точок, зварюваність матеріалу, флокочутливість та схильність до відпускнуї крихкості. Далі знаходимо таблицю з механічними властивостями, потім із першого рядка таблиці збираємо їхні назви, а з другого в циклі – можливі значення. Подібними до попереднього пункту діями знаходимо й зберігаємо таблицю з фізичними властивостями матеріалу. Отже, піс-

ля завершення роботи парсера для кожного матеріалу маємо масив його властивостей, записаний у реляційну базу даних, яку було створено для зберігання цієї інформації. Загальний алгоритм роботи парсера представлено на рис. 2,а.

Для відображення зібраних даних створено веб-застосунок, що дає змогу переглядати відомості про матеріали, виконувати пошук та інше. Можливість користувача у веб-застосунку зображено на рис. 2,б.

Сталь конструкционная углеродистая качественная 08

Применение: Сталь конструкционная углеродистая качественная

Химический состав материала:

C	Si	Mn	Ni	S	P	Cr	Cu	As
0.05 - 0.12 %	0.17 - 0.37 %	0.35 - 0.65 %	до 0.3 %	до 0.4 %	до 0.035 %	до 0.1 %	до 0.3 %	до 0.08 %

Найти похожие по хим. составу

Температура критических точек материала:
Ac1 = 735, Ac3(Acm) = 874, Ar3(Arcm) = 854, Ar1 = 680

Технологические свойства материала:
Свариваемость: без ограничений.

Флюэночувствительность: не чувствительна.

Most similar materials for Сталь конструкционная углеродистая качественная 08:

Material	Similarity index
10	0.18282505298782223
15	0.20718349355100663
12K	0.21478186608743302
15K	0.2237325412183038
Ст3сп	0.22874057357626787
20	0.22896506283710624
ВСт3сп	0.23212335082882118
20K	0.24011715890373184

Рис. 3. Пошук схожих матеріалів: а – запуск пошуку; б – результати

Після запуску функції система визначає для якого матеріалу знаходяться аналоги і починаємо визначати індекси схожості (I_s) між ним і всіма іншими матеріалами в системі за формулою

$$I_s = \sqrt{\sum_{i=1}^N (X_i^A - X_i^B)^2}, \quad (1)$$

де: N – загальна кількість хімічних елементів, що входять до складу двох матеріалів; X_i^A – концентрація i -го хімічного елементу в першому матеріалі; X_i^B – концентрація i -го хімічного елементу в другому матеріалі відповідно.

Чим менше знайдений індекс схожості, тим ближче за хімічним складом знайдено матеріал до базового (див. рис. 3, б).

Візуалізація великорозмірних наборів даних. Одним із найперших підходів до зменшення розмірності наборів даних є алгоритм Isomap (Samko, Marshall & Rosin, 2006), який підходить для ізометричного відображення. Алгоритм Isomap шукає менший розмір, який підтримує геодезичні відстані між усіма точками. Для цього Далі реалізація алгоритму Isomap складається з трьох етапів:

1. Визначаються відстані між усіма точками, з яких будується граф найближчих сусідів.
2. Пошук найкоротшого шляху в графі. Найбільш ефективними відомими алгоритмами для цього є алгоритми Дейкстра та Флойд-Варшалла.
3. Часткова декомпозиція власних значень. Вбудовування кодується у власних векторах, що відповідають d найбільшим власним значенням $N \times N$ isomap ядра.

Багатомірне масштабування (MDS, (Wickelmaier, 2003)) шукає представлення даних у низьковимірному просторі, в якому відстані можна порівняти з відстанями в початковому високовимірному просторі. Цей алгоритм використовують для аналізу даних подібності або відмінності. MDS намагається моделювати дані подібності або відмінності як відстані в геометричних просторах. Існує два типи алгоритму MDS: метричний та не-

Пошук схожих матеріалів. Досить часто виникає ситуація, коли потрібно швидко підібрати схожий за якоюсь властивістю матеріал, звичайний пошук за назвою не дає змоги це зробити з належною швидкістю. Тому в застосунку було реалізовано функцію інтелектуального пошуку аналога матеріалу за хімічним складом.

Функція доступна для кожного матеріалу в системі та запускається кліком по кнопці "Найти похожие по хим. составу" (рис. 3,а).

метричний. У метричному MDS матриця подібності вводу виникає з метрики (i в такий спосіб дотримується трикутної нерівності), тоді відстані між двома вихідними точками встановлюються якнайближче до даних подібності або невідповідності. У неметричній версії алгоритми намагаються зберегти порядок відстаней i , отже, шукати монотонні зв'язки між відстанями вбудованого простору та схожості / несхожості. Мету визначаємо за формулою

$$S = \sum_{i < j} d_{ij}(X) - \hat{d}_{ij}(X), \quad (2)$$

де: S – матриця подібності; X – координати n вхідних точок; \hat{d}_{ij} – диспропорції, перетворення подібностей, обраних оптимальними шляхами. Алгоритм t-SNE, який також відносять до методів множинного навчання ознак, опублікували в 2008 р. (Maaten, Van Der & Hinton, 2008) голландський дослідник Лоуренс ван дер Маатен і британський інформатик Джеффри Хінтон, а "класичний" SNE запропонували Хінтон і Ровейс у 2002 р. (Hinton & Roweis, 2003).

Ідея методу SNE полягає в подальшому. Є набір даних із точками, описаними багатовимірними змінними з розмірністю простору істотно більше трьох. Необхідно отримати нову змінну, яка існує у двовимірному або тривимірному просторі, яка б максимально зберігала структуру і закономірності у вихідних даних. Ідея SNE полягає у перетворенні багатовимірної евклідової дистанції між точками в умовні ймовірності, що відображають схожість точок. Загалом кроки алгоритму можна описати так:

- 1) дані x_1, \dots, x_N ;
- 2) обчислюємо p_{ij} – "схожості" x_i та x_j :

$$p_{j|i} = \frac{e^{-\|x_i - x_j\|^2 / 2\sigma_i^2}}{\sum_{k \neq i} e^{-\|x_i - x_k\|^2 / 2\sigma_i^2}}, \quad p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N}, \quad i, j = \overline{1, N}, \quad (3)$$

σ_i вибираємо, виходячи з перплексії $2^H = 2^{-\sum_x p(x) \log_2 p(x)}$;

3) знаходимо $y_1, \dots, y_N \in \mathbb{R}^d (d=2)$ у просторі меншої розмірності, схожості між якими максимально близькі до p_{ij} . Схожості для y_i та y_j знаходимо за формулою

$$q_{ji} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_i\|^2)^{-1}}, i, j = \overline{1, N}. \quad (4)$$

Візуалізації багатомірних наборів даних про хімічний склад матеріалів. Як дані є хімічні склади всіх матеріалів, які знаходимо в розмірності 118 (по одній на кожен елемент періодичної таблиці Менделєєва). То-

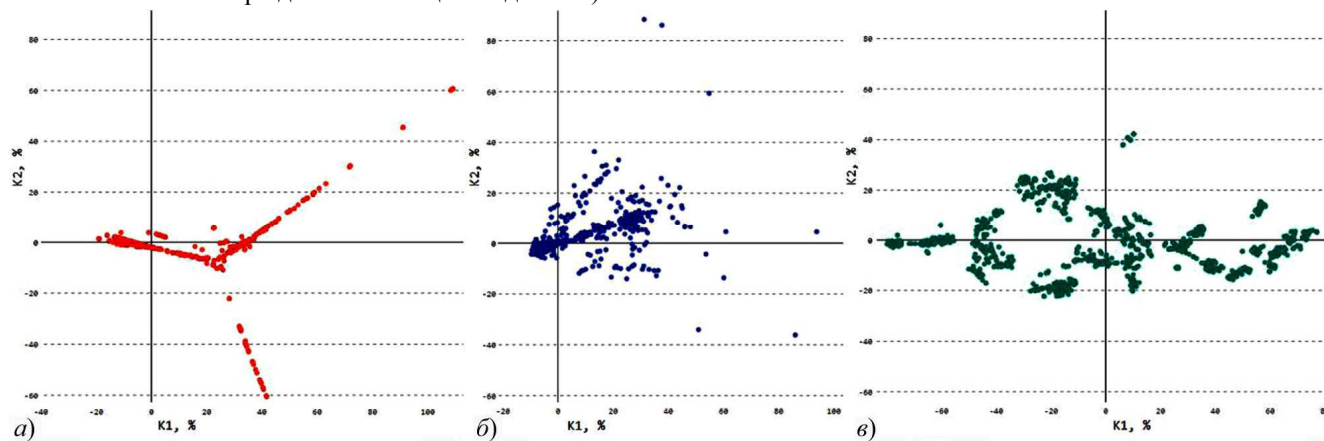


Рис. 4. Результат спрощення набору даних алгоритмами: а) Isomap; б) MDS; в) t-SNE

Як видно з рис. 4, у результатах роботи алгоритмів спостерігаємо значні відмінності. Результатом роботи алгоритму Isomap (див. рис. 4,а) спостерігаємо три яскраво виражені промені. За візуалізації методом MDS (рис. 4,б) видно локальні скупчення точок. Ці скупчення мають відмінну форму, порівняно з результатами Isomap. Однак також можна побачити три промені. Внаслідок роботи t-SNE (див. рис. 4,в) відсутня променева структура, а самі точки представлено у вигляді скупчень (кластерів). Проаналізувавши структуру кластерів, видно, що ці кластери сформовано не випадково, а вони відповідають групам матеріалу (рис. 5): сталі електротехнічні сірчисті; сталі конструкційні вуглецеві; сталі конструкційні низьколеговані; сталі жароміцні; сталі конструкційні леговані та інші.

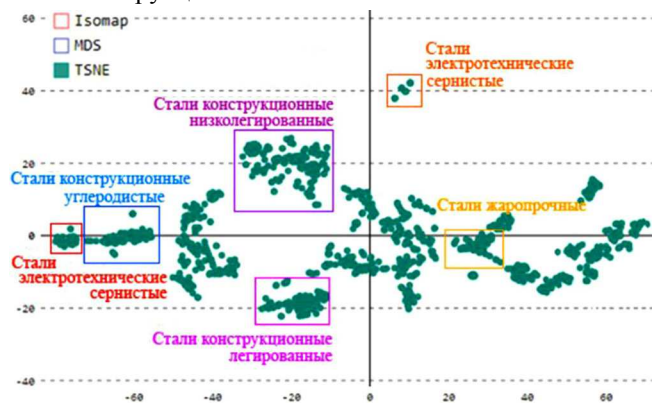


Рис. 5. Сформовані кластери матеріалів внаслідок роботи алгоритму t-SNE

вдячність. Роботу виконано за підтримки МОН України в рамках реалізації науково-дослідної роботи "Розроблення методів математичного моделювання поведінки нових та композиційних матеріалів для оцінки ресурсу та прогнозування надійності елементів конструкцій" (№ ДР 0117U004969).

Висновки. Внаслідок виконання роботи було розроблено веб-застосунок, який дає змогу в зручному ви-

му ставимо завдання щодо спрощення набору даних до розмірності 2 з метою їх відображення на площині (Samko, Marshall & Rosin, 2006).

Для зниження розмірності системи використовували готові реалізації алгоритмів бібліотеки *Scikit-learn* (Pedregosa et al., 2011; Documentation, 2018) – Isomap, MDS та t-SNE. Результати роботи алгоритмів візуально наведено на рис. 4.

ді отримати доступ до різних типів металів, виконати пошук матеріалу за назвою і підібрати схожі за хімічним складом, а також подивитися залежності між властивостями матеріалу в графічному вигляді на прикладі результатів роботи алгоритмів зниження розмірності даних.

Під час виконання роботи було розроблено й реалізовано структуру реляційної бази даних, яка дає змогу зберігати інформацію про властивості матеріалів у зручному вигляді, а також написати допоміжну програму-парсер, що значно спрощує та пришвидшує процес пошуку і відбору необхідної інформації.

Для зниження розмірності великорозмірних наборів даних було використано реалізації алгоритмів Isomap, MDS та TSNE з бібліотеки *Scikit-learn*. Внаслідок їхньої роботи матеріали було представлено точками у просторі. Найкращою візуалізацією способом зниження розмірності вдалося досягти за допомогою алгоритму t-SNE. Кластери з матеріалів, які сформовано за допомогою алгоритму, були найбільш чіткими та очевидними.

Перелік використаних джерел

- Aggarwal, S., & Gupta, S. (2014). *Flask Framework Cookbook*. Packt Publishing Ltd, 258 p.
- Diubua, P., & Sher, P. (2005). *MySQL. Sbornik receptov*. St.-Petersburg: Simvol, 1056 p. [In Russian].
- Documentation. (2018). *Manifold learning – scikit-learn 0.20.1 documentation*. Retrieved from: <https://scikit-learn.org/stable/modules/manifold.html>.
- Dwyer, G., Aggarwal, S., & Stouffer, J. (2017). *Flask: Building Python Web Services*.
- Grinberg, M. (2018). *Flask web development: developing web applications with python – miguel grinberg – Google books*.
- Hinton, G., & Roweis, S. (2003). *Stochastic neighbor embedding*.
- Jolliffe, I. (2011). *Principal component analysis. International Encyclopedia of Statistical Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1094–1096.
- Maaten, L. Van Der, & Hinton, G. (2008). *Visualizing data using t-sne*, pp. 2579–2605.
- Maia, T. (2015). *Building Web Applications with Flask*.

Pedregosa, F., Michel, V., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Samko, O., Marshall, A. D., & Rosin, P. L. (2006). Selection of the optimal parameter value for the isomap algorithm. *Pattern Recognition Letters*, 27(9), 968–979.

Wickelmaier, F. (2003). *An introduction to MDS*.

Р. Р. Лавщенко, А. А. Водка

Национальный технический университет "Харьковский политехнический институт", г. Харьков, Украина

РАЗРАБОТКА АЛГОРИТМОВ ВИЗУАЛИЗАЦИИ СВОЙСТВ ГЕТЕРОГЕННЫХ И КОМПОЗИЦИОННЫХ МАТЕРИАЛОВ

Сегодня в человеческой деятельности используется очень большое количество различных материалов. Эти материалы отличаются между собой механическими свойствами, цветом, химическим составом. Широкое разнообразие материалов ставит задачу о сохранении информации о них и про ее визуализацию. Создана база данных материалов и веб-приложение, дающие возможность получить доступ к информации о различных видах металлов и их сплавах, а также подобрать ближайшие аналоги с помощью функции интеллектуального поиска. Для создания базы данных использована система управления базами данных MySQL. Наполнение этой базы данных происходило путем разработки специальной программы-парсера. Эта программа просматривала страницы из сети интернет и выделяла информацию о свойствах материалов и сохраняла их в базу данных. Для визуализации данных использованы алгоритмы снижения размерности. Эти алгоритмы направлены на отыскание такой проекции высокоразмерных данных на низкоразмерное пространство. При использовании такой проекции сохраняются все внутренние взаимосвязи между данными. Это дало возможность построить отображение высокоразмерных данных на плоскость и визуализировать их. Для этого были использованы алгоритмы Isomap, MDS, t-SNE. Таким образом, разработана база данных материалов. Доступ к базе данных происходит с помощью веб-приложения. Для собранных данных построены визуализации. По результатам визуализации обнаружены кластеры материалов, соответствующих уже известным классам материалов. Это подтверждает корректность построения моделей визуализаций и снижения размерности, а также правильность собранной информации.

Ключевые слова: веб-приложение; металлы; алгоритмы понижения размерности; многомерные наборы данных; интеллектуальный поиск; Isomap; MDS; t-SNE.

R. R. Lavshchenko, O. O. Vodka

National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

DEVELOPMENT OF ALGORITHMS FOR VISUALIZING THE PROPERTIES OF HETEROGENEOUS AND COMPOSITE MATERIALS

Today, a lot of different materials are used in human activity. These materials are distinguished by their mechanical properties, colour, and chemical composition. A wide variety of materials raises the problem of keeping information about them and about their visualization. A database of materials and an online application that provides access to information on various types of metals and their alloys, as well as to find the nearest analogues with the help of intelligent search function is created in the work. To create the database, MySQL database management system is used. Filling of this database is done by developing a special parser program. This program scanned pages from the Internet and highlighted the properties of the materials and stored them in the database. Dimensional reduction algorithms are used to visualize the data. These algorithms are aimed at finding such a projection of high-dimensional data into low-dimensional space. When using this projection, all internal relationships between the data are stored. This made it possible to construct and visualize large-dimensional data on a plane. To do this, the isomap, mds, t-sne algorithms are used. Thus, the database of materials was developed in the work. Access to the database provided through a web application. The web application provides the ability to use it from all devices that have access to the Internet and eliminates the need to install additional software. The web application uses a python programming language and a special framework Flask for creating web applications. This technology choice greatly simplifies the creation of web applications and allows us to quickly and flexibly manage your data. Visualizations are built for the collected data. To visualize data, one of the main tasks is to reduce the dimension of data. Reducing dimensionality requires accessible visualization methods that work effectively in two or three dimensional spaces. According to the results of the visualization, clusters of materials that correspond to already known classes of materials are revealed. This confirms the correctness of constructing visualization models and reducing the dimensionality, as well as the correctness of the information gathered.

Keywords: web application; metals; manifold learning; multidimensional data sets; intelligent search; isomap; MDS; t-SNE.