В.С. Харітонов¹, Ю.Е. Парфьонов²

¹ *IT Craft, Харків*

² *Харківський національний економічний університет ім. С. Кузнеця, Харків*

АНАЛІЗ АЛГОРИТМІВ ОЦІНЮВАННЯ ВІДПОВІДЕЙ НА ТЕСТОВІ ЗАВДАННЯ ВІДКРИТОГО ТИПУ

У статті розглядається проблематика комп'ютерного оцінювання знань та навичок за допомогою тестових завдань відкритого типу. Алгоритмізація процесу оцінювання відповідей на такі завдання стикається з низкою проблем, головним чином, пов'язаних з аналізом текстів природною мовою.

Досліджено алгоритми, які базуються на методі шинглів, близькості слів з вагами та критеріальному методі з використанням елементів теорії графів. Проведено їхній порівняльний аналіз.

Ключові слова: *тестові завдання відкритого типу, семантична близькість, системи комп'ютерного тестування, комп'ютерна лінгвістика, обробка природної мови, еталонна відповідь.*

Вступ

Як показує практика, автоматизація в сфері оцінювання знань значною мірою забезпечується засобами комп'ютерного тестування. Однак, розроблення тестів і їхнє використання стикається з низкою проблем.

Зазвичай, контроль і оцінювання знань шляхом тестування зводяться до пред'явлення людині, що проходить тестування, фіксованої кількості завдань і різних варіантів відповідей на кожне з них. Їй залишається тільки вибрати один або декілька варіантів. Реалізація цього способу заснована на використанні досить простого алгоритму, який аналізує правильність чи неправильність вибраного варіанту відповіді. Основним недоліком комп'ютерного тестування з використанням тільки тестових завдань закритого типу є те, що перевірка глибинного розуміння предмета, оволодіння стилем мислення, притаманним певній дисципліні, часто ускладнена.

Можливим варіантом подолання цього недоліку є використання тестових завдань відкритого типу. Проте, оцінювання відповідей на такі завдання має особливу складність, оскільки для аналізу фраз природною мовою потрібні більш досконалі алгоритми.

У загальному випадку, аналіз фраз природною мовою являє собою дуже актуальну проблему, особливо останнім часом, з причини великого зростання обсягів текстової інформації. На сьогоднішній день не існує ні точного алгоритму, ні автоматизованої системи тестування, що задовольняє всім вимогам. Багато проблем виникає саме під час алгоритмізації оцінювання відповідей на тестові завдання відкри-

того типу. Існують різні методи та алгоритми, але кожен алгоритм оцінює одну й ту саму відповідь по різному.

В даній статті розглядаються три алгоритми.

Перший алгоритм оснований на методі шинглів. Ключовим моментом алгоритму є порівняння відповідей, отриманих від людини, що відповідає на запитання, та еталонних. Для цього пропонується використовувати апарат виявлення схожості текстів, що базується на формуванні і наступному порівнянні відповідних шинглів. Більш детально даний алгоритм розглянуто в роботах [1–3].

Другий алгоритм оснований на близькості слів та їхніх ваг. У цьому алгоритмі центральним об'єктом аналізу стає не весь текст відповіді, а його речення. Речення – це мінімальна одиниця мови, яка представляє собою граматично організоване з'єднання слів (або слово), що володіє смисловою завершеністю [4–9].

Останній алгоритм базується на критеріальному методі з використанням елементів теорії графів. Еталонна модель є основою для когнітивно-комунікативного підходу оцінювання відповідей на запитання відкритого типу. В даному алгоритмі роль еталона виконує лексико-семантична модель конкретної предметної області. Для формального опису такої моделі використовується модифікований ідеографічний метод Караулова Ю.М. з використанням елементів теорії графів [10–14].

Для вирішення питань, які пов'язані з вибором алгоритму оцінювання відповідей на тестові завдання відкритого типу та його ефективністю, слід провести порівняльний аналіз. Це і є головною метою даної роботи.

Виклад основного матеріалу

Таким чином, для проведення порівняльного аналізу було вибрано три алгоритми. Головною причиною вибору саме цих алгоритмів є те, що в їх реалізації застосовуються різні підходи, основні риси яких наведено далі.

Алгоритм, який базується на методі шинглів:

- використовується для перевірки текстів на унікальність;
- потребує канонізації тексту;
- вхідними даними є лише еталонний текст та відповідь;

– центральним об'єктом виступає весь текст.

Алгоритм на основі близькості слів з вагами:

- враховує важливість слова відносно інших слів;

- потребує канонізації тексту;

- вхідними даними є еталонний текст, відповідь та ключові слова з їх вагами;

- центральним об'єктом виступають речення.

Алгоритм, який реалізує критеріальний метод з використанням елементів теорії графів:

- враховує предметну область знань та критерії;

- вхідними даними є еталонна предметна область, яка має графову структуру, та відповідь;

- центральним об'єктом виступає предметна область.

Для проведення порівняльного аналізу наведених алгоритмів була розроблена автоматизована система тестування.

З урахуванням класифікацій запитань та відповідей було підготовлено 5 запитань з еталонами відповідей. Дані запитання мають різну конструкцію та тип, для того, щоб була можливість оцінити ефективність роботи алгоритмів.

Запитання 1. Наведіть значення терміну “Клас” (термін).

Еталон 1: Спеціальна конструкція, яка використовується для групування пов'язаних змінних та функцій (предметна область – програмування).

Еталон 2: Кімната в школі, гімназії тощо, де здійснюється навчальний процес (предметна область – навчання).

Ваги 1: “клас” – 5, “спеціальна” – 3, “конструкція” – 8, “групування” – 7, “змінних” – 6, “функцій” – 6.

Ваги 2: “клас” – 5, “здійснюється” – 6, “навчальний” – 8, “процес” – 8, “кімната” – 10, “школі” – 5, “гімназії” – 5.

Предметні області 1, 2 та 3 зображені на рис. 1–3 відповідно.

Відповідь 1: Для групування пов'язаних функцій та змінних використовується спеціальна конструкція – клас (відповідь з перестановкою слів).

Відповідь 2: Клас – це спеціальна конструкція (неповна відповідь).

Відповідь 3: Клас – кімната в школі, де здійснюється навчальний процес (неправильна відповідь для першої та третьої предметних областей, але правильна для другої).

Рис. 1. Предметна область 1 (запитання 1)

Рис. 2. Предметна область 2 (запитання 1)

Рис. 3. Предметна область 3 (запитання 1)

Запитання 2. Як називається браузер для IOS?
(запитання з точною короткою відповіддю).

Еталон: Safari.

Ваги: “Safari” – 10.

Предметна область зображена на рис. 4.

Рис. 4. Предметна область (запитання 2)

Відповідь 1: Safari (правильна відповідь).

Відповідь 2: Chrome (неправильна відповідь).

Відповідь 3: Браузер для IOS – це Safari (правильна, але не коротка відповідь).

Запитання 3. Перерахуйте основні типи даних, які існують в JavaScript (перерахування).

Еталон: number, string, boolean, null, undefined, object.

Ваги: “number” – 7, “string” – 7, “boolean” – 7, “null” – 10, “undefined” – 10, “object” – 10.

Предметна область зображена на рис. 5.

Рис. 5. Предметна область (запитання 3)

Відповідь 1: object, number, null, boolean, undefined, string (правильна відповідь, але в іншій послідовності).

Відповідь 2: object, number, null (наполовину правильна відповідь).

Запитання 4. Дайте характеристику типу даних TIMESTAMP (запитання з розгорнутою відповіддю).

Еталон: TIMESTAMP – тип даних для зберігання дати і часу. Дані зберігаються у вигляді кількості секунд, що пройшли з початку “епохи Юнікса”. Діапазон значень: 1970/01/01 00:00:00 – 2038/12/31 00:00:00. Займає 4 байти.

Ваги: “TIMESTAMP” – 7, “тип даних” – 8, “дата” – 8, “час” – 8, “кількість секунд” – 6, “епоха Юнікса” – 3, “1970/01/01 00:00:00 – 2038/12/31 00:00:00” – 7, “4 байти” – 8.

Предметна область зображена на рис. 6.

Рис. 6. Предметна область (запитання 4)

Відповідь: TIMESTAMP – тип даних, займає 4 байти, має діапазон значень 1970/01/01 00:00:00 – 2038/12/31 00:00:00, використовується для зберігання часту і дати (правильна відповідь, але має іншу структуру характеристики).

Запитання 5. Напишіть свої думки на тему “ринок праці IT-сфери” (індивідуальне розгорнуте питання).

Еталон: Щодня з’являються нові гаджети, що вимагає дедалі більшої кількості програмістів. Смартфони, планшети, окуляри доповненої реальності, гіроскутери, дрони, безпілотні автомобілі, фітнес-браслети, смарт-годинники – все це потребує величезної армії програмістів, які щодня створюють та оновлюють застосунки, щоб спростити різні процеси повсякденного життя. Автоматизація виробництва й офісної роботи розвивається з кожним днем. На сьогодні більшість працівників виробничих компаній замінили роботи, а вже незабаром, як вважають, цілі відділи в корпораціях зможе замінювати програміст, який буде обслуговувати нейромережу, що сама здатна приймати рішення, навчатися та вдосконалюватися. Сьогодні в багатьох країнах програмування вчать із самого дитинства, і для цього спеціально розроблені навчальні мови. Передбачають, що через декілька років знання основ програмування стане невід’ємною частиною будь-якої освіченої людини, чи то вона менеджер, чи то юрист, чи то агроном.

Ваги: “нові гаджети” – 7, “автоматизація” – 9, “застосунки” – 7, “спрощення” – 8, “розвиток” – 9, “програміст” – 6, “нейромережа” – 3, “інформаційні технології” – 6, “умови праці” – 8.

Предметна область зображена на рис. 7.

Рис. 7. Предметна область (запитання 5)

Відповідь: Інформаційні технології – одна з небагатьох професійних сфер, яка найменше постраждала за часів кризи.

Оскільки ІТ-галузь в більшій мірі орієнтована не на внутрішній, а на світовий ринок, робота з іноземними замовниками дає можливість не тільки платити високі зарплати своїм співробітникам, але і пропонувати хороші умови праці. Також, за даними опитування, представники українських компаній вважають розвиток інформаційних технологій найбільш стратегічно важливим напрямком для України поряд з агробізнесом, сільським господарством і виробництвом.

В розробленій автоматизованій системі тестування алгоритм, який оснований на методі шинглів, називається “Shingles”, алгоритм на основі близькості слів з вагами називається “Proximity Of Words”, алгоритм, який реалізує критеріальний метод з використанням елементів теорії графів, називається “Graph”. Після проходження 5 тестів з різними входними даними побудовано графіки результатів оцінювання, які зображено на рис. 8.

Рис. 8. Результати оцінювання тестових завдань відкритого типу

Тут координатна вісь “Оцінки” має позначки у форматі “цифра-цифра-цифра”. Перша цифра означає номер запитання, друга – відповідь на це запитання, а третя – вибрану предметну область.

Аналізуючи одержані результати, можна сказати, що алгоритм “Shingles” досить погано справляється з відповідями, які є правильними, але мають інший порядок слів, ніж визначений еталон. Також цей алгоритм не вміє аналізувати синоніми. Але він непогано оцінює відповіді в короткій формі або ті, де порядок слів дійсно важливий. Це особливо помі-

тно на показнику “2-1-1”.

Дуже відрізняється оцінки “1-3-2”. Це демонструє унікальну особливість алгоритму “Graph”. Є запитання, на які відповідь може бути кардинально різною, тому що залежить від контексту або предметної області. Алгоритми “Shingles” і “Proximity of Words” можуть оцінити відповідь тільки в рамках еталонної відповіді, що не може вийти за рамки однієї предметної області, а охоплення алгоритму “Graph” нічим не обмежуються. Але, завдяки цьому є і недоліки. Алгоритм “Graph” погано оцінює дуже короткі

відповіді, де правильною оцінкою є 0 або 100. Це видно з показників “2-1-1”, “3-1-1” та “5-1-1”. Цей алгоритм є складним, як у реалізації, так і у використанні, тому що для його точної роботи слід побудувати складну графову структуру. Також слід зазначити, що всі алгоритми, за винятком “Graph”, зовсім не впоралися з показником “5-1-1”. Це дуже складний тип запитань або мислення для всіх автоматизованих систем, оскільки йде мова про творчу відповідь та персональні думки. Алгоритм “Proximity of Words” показує кращі показники ніж “Shingles” та конкурує з “Graph”. Він має середню складність реалізації та потребує додатково лише ключові слова з вагами.

Висновки

Розглянуто проблему комп'ютерного тестування знань та навичок з використанням тестових завдань відкритого типу. Автоматизація оцінювання

відповідей на такі завдання є складною проблемою. В рамках даної роботи було досліджено алгоритм на базі методу шинглів, алгоритм на основі близькості слів з вагами та алгоритм на базі критеріального методу з використанням елементів теорії графів. Кожен з цих алгоритмів виконує одну задачу – аналіз текстів. Але між ними є фундаментальні відмінності у тому, як вони це роблять. Для з'ясування цих відмінностей було проведено порівняльний аналіз з використанням розробленої автоматизованої системи тестування. Він дав змогу з'ясувати, що алгоритмізація оцінювання відповідей на тестові завдання відкритого типу можлива, але немає унікального алгоритму на всі випадки. Кожен з алгоритмів має як сильні, так і слабкі сторони. За результатами аналізу найперспективнішим став алгоритм на базі критеріального методу з використанням елементів теорії графів.

Список літератури

1. Официальный сайт сообщества IT-специалистов Хабр. Поиск нечетких дубликатов [Электронный ресурс]. – Режим доступа: <https://habrahabr.ru/post/65944>.
2. Негода В.Н. Информатика и вычислительная техника / В.Н. Негода. – Ульяновск: Изд. УлГТУ, 2010. – 677 с.
3. Атанов Г.А. Обучение и искусственный интеллект, или основы современной дидактики высшей школы / Г.А. Атанов, И.Н. Пустышников. – Донецк: Изд. ДООУ, 2002. – 504 с.
4. Марков В.В. Организация тестового контроля знаний на основе нечеткой модели экзаменатора / В.В. Марков, М.В. Луцан // Известия ЮФУ. Технические науки. – 2010. – № 7(108). – С. 242-248.
5. Технологии интеллектуального анализа и извлечения данных на основе принципов эволюционного моделирования / В.В. Бова, Л.А. Гладков, Ю.А. Кравченко, В.В. Курейчик, В.М. Курейчик, С.Н. Щеглов. – Таганрог: Изд. ТТИ ЮФУ, 2009. – 124 с.
6. Лукашевич Н.В. Тезаурусы в задачах информационного поиска / Н.В. Лукашевич. – М.: Изд. Московского университета, 2011. – 512 с.
7. Савинов А.П. Интеллектуализация обучающей системы Moodle позволит продлить срок эксплуатации ее в вузах / А.П. Савинов, Т.С. Петровская, Д.И. Фирстов // Высшее образование сегодня. – 2014. – № 9. – С. 15-21.
8. Двухуровневый лингвистический процессор ответных текстов на естественном языке / Д.Ш. Сулейманов. – Минск: Изд. OSTIS, 2011. – С. 311-322.
9. Burrows S., Gurevych I., Stein B. The eras and trends of automatic short answer grading / S. Burrows, I. Gurevych, B. Stein // International Journal of Artificial Intelligence in Education, - 2015. – Volume 25, Issue 1. - p. 60-117.
10. Краснов А.Н. Психолого-педагогические основы технологии открытого тестирования / А.Н. Краснов. – Самара: Изд. СТИ, 2005. – С. 348-399.
11. Андрусенко Т.Б. Лингвистические структуры в компьютерных учебных средах / Т.Б. Андрусенко; НАН Украины, Ин-т кибернетики им. В. М. Глушкова. – Киев: Наук. думка, 1994. – 160 с.
12. Евстигнеев В.А. Теория графов. Алгоритмы обработки деревьев / В.А. Евстигнеев, В.Н. Касьянов. – Новосибирск: Наука, 1994. – 360 с.
13. Караулов Ю.Н. Общая и русская идеография / Ю.Н. Караулов. – М.: Либроком, 2016. – 356 с.
14. Клековкин Г.А. Геометрическая теория графов / Г.А. Клековкин, Л.П. Коннова, В.В. Коннов. – М.: ЮРАЙТ, 2017. – 240 с.

References

1. The official site of IT specialists' community Khabr (2020), “*Poisk nechetkikh dublikatov*” [Search for fuzzy duplicates], available at: <https://habrahabr.ru/post/65944> (accessed 10 January 2020).
2. Negoda, V.N. (2010), “*Informatika i vychislitel'naya tekhnika*” [Computer Science and Computer Engineering], UIGTU, Ulyanovsk, 677 p.
3. Atanov, G.A. and Pustynnikova, I.N. (2002), “*Obucheniye i iskusstvennyy intellekt, ili osnovy sovremennoy didaktiki vysshey shkoly*” [Training and artificial intelligence, or the foundations of modern higher education didactics], DOU, Donetsk, 504 p.
4. Markov, V.V. and Lutsan, M.V. (2010), “*Organizatsiya testovogo kontrolya znaniy na osnove nechetkoy modeli ekzamenatora*” [Organization of test control of knowledge based on a fuzzy model of the examiner], *Izvestiya YUFU. Tekhnicheskiye nauki*, No. 7(108), pp. 242-248.
5. Bova, V.V., Gladkov, L.A., Kravchenko, Y.A., Kureychik, V.V., Kureychik, V.M. and Shcheglov, S.N. (2009), “*Tekhnologii intellektual'nogo analiza i izvlecheniya dannykh na osnove printsipov evolyutsionnogo modelirovaniya*” [Data mining technologies based on principles of evolutionary modeling], Izd. TTI YUFU, Taganrog, 124 p.

6. Lukashevich, N.V. (2011), "Tezaurusy v zadachakh informatsionnogo poiska" [Thesauri in information retrieval problems], Izd. Moskovskogo universiteta, Moscow, 512 p.
7. Savinov, A.P., Petrovskaya, T.S. and Firstov, D.I. (2014), "Intellectualizatsiya obuchayushchey sistema Moodle pozvolyt prodlit srok ekspluatatsii yeye v vuzakh" [Intellectualization of the Moodle training system will extend its life in universities], Vyssheye obrazovaniye segodnya, No. 9, pp. 15-21.
8. Suleymanov, D.Sh. (2011), "Dvukhurovnevyy lingvisticheskiy protsessor otvetnykh tekstov na yestestvennom yazyke" [Two-level linguistic natural language response processor], Izd. OSTIS, Minsk, pp. 311-322.
9. Burrows, S., Gurevych, I. And Stein, B. (2015), The eras and trends of automatic short answer grading, International Journal of Artificial Intelligence in Education, Volume 25, Issue 1, pp. 60-117.
10. Krasnov, A.N. (2005), "Psikhologo-pedagogicheskiye osnovy tekhnologii otkrytogo testirovaniya" [Psychological and pedagogical foundations of open testing technology], Izd. STI, Samara, pp. 348-399.
11. Andrusenko, T.V. (1994), "Lingvisticheskiye struktury v kompyuternykh uchebnykh sredakh" [Linguistic structures in computer learning environments], Nauk. dumka, Kyiv, 160 p.
12. Yevstigneyev, V.A. and Kasyanov, V.N. (1994), "Teoriya grafov. Algoritmy obrabotki derevyev" [Graph theory. Tree Processing Algorithms], Nauka, Novosibirsk, 360 p.
13. Karaulov, Y.N. (2016), "Obshchaya i russkaya ideografiya" [General and Russian ideography], Librokom, Moscow, 356 p.
14. Klekovkin, G.A., Konnova, L.P. and Konnov, V.V. (2017), "Geometricheskaya teoriya grafov" [Geometric graph theory], YURAYT, Moscow, 240 p.

Надійшла до редколегії 05.02.2020

Схвалена до друку 10.03.2020

Відомості про авторів:**Харітонов Владислав Станіславович**

Full Stack JavaScript-розробник
IT Craft,
Харків, Україна
<https://orcid.org/0000-0002-1199-8350>

Парфонов Юрій Едуардович

кандидат технічних наук старший науковий співробітник
доцент кафедри
Харківського національного економічного
університету ім. С. Кузнеця,
Харків, Україна
<https://orcid.org/0000-0003-3943-3201>

Information about the authors:**Vladyslav Kharitonov**

Full Stack JavaScript Software Engineer
of IT Craft,
Kharkiv, Ukraine
<https://orcid.org/0000-0002-1199-8350>

Yurii Parfonov

Candidate of Technical Sciences Senior Research
Senior Lecturer
of Simon Kuznets Kharkiv National
University of Economics,
Kharkiv, Ukraine
<https://orcid.org/0000-0003-3943-3201>

АНАЛИЗ АЛГОРИТМОВ ОЦЕНИВАНИЯ ОТВЕТОВ НА ТЕСТОВЫЕ ЗАДАНИЯ ОТКРЫТОГО ТИПА

В.С. Харитонов, Ю.Э. Парфенов

В статье рассматривается проблематика компьютерного оценивания знаний и навыков с помощью тестовых заданий открытого типа. Алгоритмизация процесса оценки ответов на такие задания сталкивается с рядом проблем, главным образом, связанных с анализом текстов на естественном языке.

Исследованы алгоритмы, основанные на методе шинглов, близости слов с весами и критериальном методе с использованием элементов теории графов. Проведен их сравнительный анализ.

Ключевые слова: тестовые задания открытого типа, семантическая близость, системы компьютерного тестирования, компьютерная лингвистика, обработка естественного языка, эталонный ответ.

ANALYSIS OF ALGORITHMS TO ASSESS ANSWERS ON OPEN-ENDED TEST TASKS

V. Kharitonov, Yu. Parfonov

The article considered the problem of computer control of knowledge and skills. One of the options of such control is computer testing. Usually, the control and assessment of knowledge through testing are reduced to presenting the person undergoing testing, a fixed number of tasks and different answers to each of them. He (or she) only should choose one or more options. The implementation of this method is based on the use of a fairly simple algorithm that analyzes the correctness or incorrectness of the chosen answer. The main disadvantage of computer-based testing using only closed-ended test tasks is that validation of a deep understanding of the domain, mastery of a style of thinking inherent in a discipline, is often complicated. A possible way to overcome that drawback is by using open-ended test tasks. However, automating the assessment of answers to such tasks faces many problems, mainly related to the analysis of natural language texts. There are different methods and algorithms, but each algorithm evaluates the same answer differently. In the work have researched algorithms based on the method of w-shingles, on the proximity of words with weights and the criterion method using elements of graph theory. Each of these algorithms performs the same task - analysis text in natural language. But there are fundamental differences between them in how they do it. To clarify these differences, a comparative analysis was conducted using developed automated testing system. It made it clear that algorithmizing the evaluation of answers to open-ended test questions is possible, but there is no unique algorithm for all cases. Each of the algorithms has both strengths and weaknesses. According to the results of the analysis, the most promising is an algorithm based on the criterion method using graph theory. The results of the work can be used to further develop the evaluation algorithms related to answers on open-ended test tasks.

Keywords: open-ended test task, semantic proximity, computer testing systems, computer linguistics, natural language processing, reference answer.