

Усунення мовної неузгодженості в тестових завданнях, згенерованих на основі понятійно-тезисної моделі.

*Танченко С. С., к.т.н., доц. Титенко С. В., к.т.н., доц. Гагарін О.О.
Національний технічний університет України «КПІ», м. Київ
serg.tanchenko@gmail.com*

У роботі запропоновано розширення понятійно-тезисної моделі (ПТМ) з метою подолання проблеми мовної неузгодженості в тестових завданнях, що будуються автоматично. Запропоновано використання узагальненого мовного класифікатора як додаткової характеристики структурних елементів ПТМ. Удосконалено алгоритм побудови тестових завдань на базі ПТМ.

Вступ

В умовах стрімкого розвитку технологій та неспішною зміною університетських програм, значну частину професійних знань фахівець отримує не в межах навчального закладу. Із підвищенням ступеня інформатизації суспільства та освіти проблема ефективного контролю знань набуває особливої значущості. Однією з найбільш поширених форм перевірки знань є комп'ютерне тестування. У попередній роботі [1] було здійснено огляд методів генерації тестових завдань та обґрунтовано ПТМ як доцільний напрямок з точки зору реалізації та практичного використання. Завданням даної роботи є розширення семантики ПТМ для усунення проблеми мовної неузгодженості тестових завдань, що часом виникають при генерації тестів.

Проблема мовної неузгодженості в тестових завданнях

Побудова тестових завдань на основі ПТМ відбувається шляхом вибору із сукупності понять та тез певного набору елементів, що стають структурними одиницями тестового завдання [2].

Розглянемо проблему мовної неузгодженості тестів, згенерованих на основі ПТМ. Проблема полягає у невідповідності роду або числа понять та тез, що беруть участь у конструюванні тестового завдання. Це суттєво знижує якість завдання, а також ефективність перевірки засвоєння

матеріалу. Нижче наведені приклади згенерованих завдань із мовними неузгодженостями:

1. *У чому полягає сутність поняття **session.name**?*
 - a) **являють** собою розширення протоколу HTTP;
 - b) повертає ім'я поточної сесії;
 - c) є інформаційною моделлю предметної області;
 - d) являє собою рядок з 32 символів.

2. *До якого поняття наступне твердження підходить найбільше?
Найбільш очевидно **вона** проявляє себе в систематичному переповненні каналів грошового обігу масою надлишкових грошей, що веде до їх знецінення:*
 - a) конкретно-економічні відносини;
 - b) пропонування;
 - c) **інфляція**
 - d) засіб обігу.

3. *У чому полягає сутність поняття: **Internet**?*
 - a) це система, що безперервно розвивається людьми, які використовують її послуги;
 - b) ідентифікує та встановлює наявність передбачуваних партнерів для зв'язку;
 - c) **дозволяла** будь-якому з комп'ютерів зв'язуватись з будь-яким іншим;
 - d) представлений протоколом TCP і UDP;
 - e) можна розділити на мережі з комутацією потоків і мережі з комутацією каналів.

4. *Яка особливість поняття: **канальний рівень**?*
 - a) нижче цього рівня тільки апаратний рівень;
 - b) **охопила** більше ніж 100 країн, **об'єднала** більше 40 тис. окремих мереж;
 - c) його політика припускала, що нова мережа буде орієнтована на різноманітних користувачів;
 - d) у 1986 році **зв'язувала** менше 6000 тис. комп'ютерів.
 - e) використовує процес експертної оцінки документів.

У тесті №2 є очевидним, що лише один варіант відповіді підходить за синтаксисом.

Вирішення проблеми мовної неузгодженості в тестових завданнях, згенерованих на основі ПТМ

Генерація тестів з використанням ПТМ відбувається на основі понятійно-тезисної бази, структурними одиницями якої є: «поняття» та «тези». Позначимо множину понять як:

$$C = \{c_1, \dots, c_{n1}\}$$

Тези – це відомості про поняття. Зазвичай теза – це одне або декілька речень, в яких мова йде про поняття, але словесно поняття не фігурує в ній. Позначимо множину тез як:

$$T = \{t_1, \dots, t_{n2}\}$$

У свою чергу кожне поняття може мати будь-яку кількість тез. Зв'язок між поняттями та тезами:

$$CT: T \rightarrow C, TC: C \rightarrow 2^T$$

ПТ-елементи виділяються із тексту навчального фрагменту, що відноситься до певного предмету. Позначимо множину предметів як:

$$S = \{s_1, s_2, \dots, s_{n4}\}$$

Зв'язок між предметами та тезами: $TS: S \rightarrow 2^T$. Відповідно до цього, поняття предмету визначаються як:

$$CS(s) = \{c | TS(s) \cap CT(c) \neq \emptyset\}$$

Шляхом подолання проблеми мовної неузгодженості є введення додаткової характеристики в опис структурних одиниць ПТМ, що отримала назву «узагальнений мовний класифікатор» (УМК). Множина УМК для поняття:

$$MC: \{mc_1, \dots, mc_{n3}\},$$

де mc_1 – «чоловічий»; mc_2 – «жіночий»; mc_3 – «середній»; mc_4 – «множина». Для кожного поняття та тези додається свій УМК. Особливістю є те, що теза може лексично відповідати поняттям які мають відмінне значення УМК, ніж сама теза. Зв'язок між поняттям і УМК: $MCC: C \rightarrow MC$. Позначимо множину УМК для тез:

$$MT = MC \cup mt',$$

де $mt' = (\{mc_1, mc_2, mc_3\})$ і означає «довільну одиницю». Зв'язок між тезою і УМК як: $MTT: T \rightarrow MT$.

На рис. 1 зображено тестове завдання для поняття «константа» – жіночого роду, проте в якості дистракторів виступають тези як жіночого так і чоловічого родів, наприклад теза поняття «клас».

Поняття	3. До якого твердження наступне поняття підходить найбільше? - константа
функція	<input type="radio"/> не повинна бути визначена до її використання
змінна	<input type="radio"/> в PHP представлена знаком долара з послідовним ім'ям
константа	<input type="radio"/> ідентифікатор простого значення
клас	<input type="radio"/> може успадковувати методи і властивості
PHP	<input type="radio"/> інтерпретується веб-сервером в HTML-код, який передається на сторону клієнта

Рис. 1. Згенероване тестове завдання

Розроблено **алгоритм побудови тестового завдання** $task_k$ на основі ПТ-елементів, що містить наступні кроки:

1) пошук контрольного поняття c_k на основі якого буде згенероване тестове завдання:

- вибір випадкового поняття c'_k з контексту контрольного предмету s_k : $c'_k: [c'_k \in C \wedge c'_k \in CS(s_k)]$;
- пошук множини альтернативних варіантів відповідей (тез) T'_k , що не відносяться до поняття c'_k :

$$T'_k = \{t | (t, c'_k) \notin CT \wedge [MTT(t) = MCC(c'_k) \vee MTT(t) \in MCC(c'_k)]\}$$

- перевірка, чи задовольняє обране поняття вимозі $|T'_k| \geq 4$, тобто контрольним поняттям може бути тільки поняття, ще має 4 або більше альтернативних варіантів відповідей. Якщо умова виконується, прийняти $c_k = c'_k$, інакше повернутись до п. 1.а;

2) вибір контрольної тези: $t_k: t_k \in T'_k$, де T'_k – множина тез, що відносяться до поняття c_k : $T_k = \{t | (t, c_k) \in CT\}$.

3) візуалізація тестового завдання.

В результаті тестове завдання складається із: $Tasks(task_k) = \langle c_k, t_k, T'_k \rangle$, де: c_k - контрольне поняття; t_k - контрольна теза; T'_k - альтернативні варіанти відповідей.

На рис. 2 показано алгоритм побудови тестового завдання.

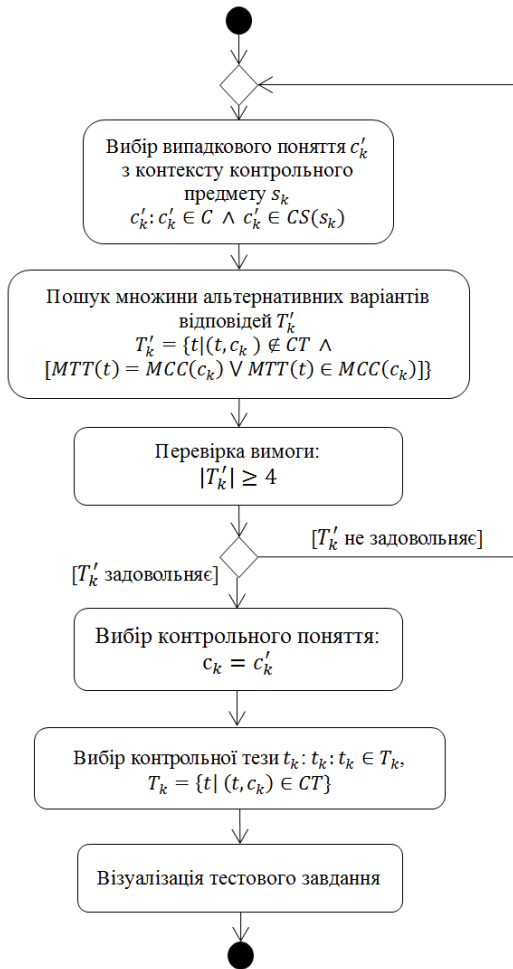


Рис. 2. Діаграма діяльності у нотатції UML “Алгоритм побудови тестового завдання”

Висновки

В роботі було описано метод подолання мовної неузгодженості згенерованих завдань. Суть методу полягає в розширенні онтології на основі понятійно-тезисної моделі шляхом введення в опис структурних

одиниць ПТМ характеристики узагальненого мовного ідентифікатора. Удосконалено алгоритм генерації тестового завдання з урахуванням вимог лексичної узгодженості понятійно-тезисних елементів.

Література

1. *Танченко С.С., Титенко С.В.* Аналіз методів генерації тестових завдань // Международная научная конференция имени Т.А. Таран «Интеллектуальный анализ информации ИАИ-2013», Киев, 15-27 мая 2013 г.: сб. тр./ ред. кол.: С.В. Сирота (гл. ред.) и др. – К.: Просвіта, 2013. – С. 218-225.
2. *Титенко С. В.* Генерація тестових завдань у системі дистанційного навчання на основі моделі формалізації дидактичного тексту / С. В. Титенко // Наукові вісті НТУУ "КПІ". – 2009. – № 1(63). – С. 47–57.