

сути «цифровой экономики», а именно, структуры и функционирования компьютеризированных систем управления производством. Поэтому, чтобы оставаться в тренде в современной экономике необходимо думать о подготовке специалистов еще в вузе, так как применение в ходе хозяйственной деятельности предприятия автоматизированных адаптивных систем управления дает возможность существенно повысить эффективность управления производством и конкурентоспособностью предприятия в целом.

Список литературы

1. Урманцева А. Термин «Цифровая экономика» как понимают современные ученые!!! / А.Урманцева. [Электронный ресурс] – <https://ria.ru/science/20170616/1496663946.html>
2. Лещенко Е. В. Адаптивная система оценки конкурентоспособности предприятия при оперативно-тактическом управлении / Е.В. Лещенко // Вдосконалення економіки та фінансової системи країни: актуальні проблеми та перспективи: матеріали доповідей Міжнародної науково-практичної конференції. – Запоріжжя : Класичний приватний університет, 2018. – Ч.2. – С. 103-105.

АВТОМАТИЧНА ЕКСТРАКЦІЯ ТЕРМІНОЛОГІЧНОЇ ІНФОРМАЦІЇ З АНГЛОМОВНИХ ТЕКСТІВ У ГАЛУЗІ SOFTWARE TESTING

Галкіна Я. Р., Петрасова С. В.

*Національний технічний університет
«Харківський політехнічний інститут»,
Харків, вул. Пушкінська, 79/2, тел. 707-63-60,
e-mail: galkina230797@gmail.com, svetapetrasova@gmail.com*

Добування інформації є одним з основних завдань автоматичної обробки мови – загального напрямку інформатики, штучного інтелекту та математичної лінгвістики, який вивчає проблеми комп'ютерного аналізу та синтезу природної мови [1].

В свою чергу до задач добування інформації входять розпізнавання іменованих сутностей, вирішення анафори і конференцій, екстракція термінології, виявлення в текстах смислової, оціночної інформації та ін.

Завдання екстракції термінів (ключових слів і фраз) з тексту виникає в багатьох областях: інформаційний пошук, електронний документообіг, моніторинг бізнес-процесів та ін. Обсяги та динаміка інформації в цих областях роблять актуальною задачу автоматичного виділення ключових слів і фраз, які можуть використовуватися для створення і вдосконалення термінологічних ресурсів, а також для ефективної обробки документів в інформаційно-пошукових системах (індексування, реферування і класифікації).

У процесі дослідження проблеми добування термінологічної інформації досягнуто значних результатів за рахунок застосування підходів, заснованих на правилах (rule-based) та використанні словникових ресурсів, а також методів машинного навчання (machine learning).

Однак, існує проблема, яка досі не знайшла свого вирішення, що коріниться у самій природі людської мови та полягає в її неоднозначності (синтаксичної, смислової, відмінкової або референційної).

В роботі з метою поліпшення якості автоматичної обробки природної мови у галузі тестування програмного забезпечення, зокрема вирішення проблеми смислової неоднозначності шуканого матеріалу, пропонується варіант побудови автоматичної системи екстракції термінологічної інформації з неструктурованого ресурсу.

Проведений аналіз текстів з навчання тестуванню програмного забезпечення (software testing) показав, що найбільш типовими ознаками є насиченість тексту спеціальними термінами та стилістична нейтральність. Однак комп'ютерна лексика англійської мови має особливості. Одна з них – притаманність окремим термінам образності.

З огляду на походження та функціонування комп'ютерна терміносистема складається із термінів, які можна поділити на такі групи: корелятивні терміни; загальнотехнічні терміни; спеціальні терміни; терміни, що мають два і більше значень у комп'ютерній галузі [2].

В даному дослідженні для автоматичної екстракції термінологічної інформації зі спеціалізованого корпусу текстів пропонується застосувати розроблені формальні шаблони розпізнавальних конструкцій та лінгвістичні правила їх обробки, а також словниковий підхід на основі розробленої бази даних термінів та їх визначень.

В результаті дослідження було розроблено та програмно реалізовано алгоритм автоматичного визначення термінів в англійських текстах з навчання тестуванню програмного забезпечення (рис. 1).

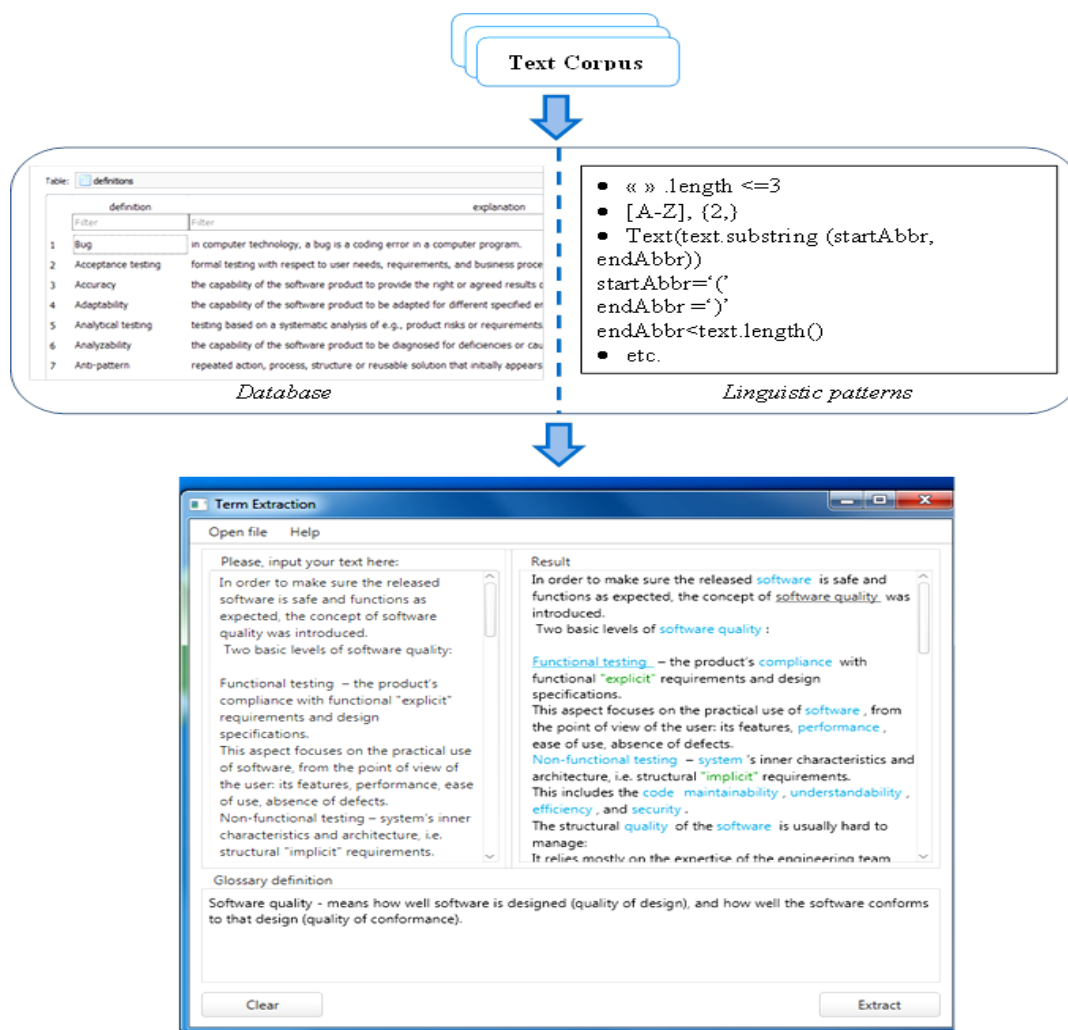


Рис. 1. Інформаційно-лінгвістичне забезпечення

Розроблена програмна імплементація може використовуватися при навчанні тестуванню програмного забезпечення, оскільки виділення термінів полегшує роботу з текстом та надає можливість одразу дізнатися значення терміну. Крім того, дана розробка може бути використана в якості лінгвістичної бази для системи тестування, завдяки якій вирішується проблема неоднозначності відповідей на природній мові.

Список літератури

1. Большакова Е.И. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика / Е.И. Большакова, Э.С. Клышинский и др. – М. : МИЭМ, 2011. – 272 с.
2. Синдега Р.Є. Структурні особливості та функціонування термінів в англomовних текстах з проблем комп'ютерних наук та інформаційних технологій / Р.Є. Синдега, О.М. Іващишин // Наукові записки. Серія «Філологічна». – Острого: Видавництво Національного університету «Острозька академія». – Вип.11. – 2009. – С. 351–357.