

ЗАСТОСУВАННЯ ПОНЯТТЯ «БРИТВА ОККАМА» ДО ЗАДАЧ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ

Вінницький національний технічний університет

Анотація

Розглянуто випадки можливого використання поняття «Бритва Оккама» для аналізу та синтезу природно-мовних конструкцій. Запропоновано формалізувати поняття бритви Оккама з метою застосування його як критерія для розв'язання визначеного класу задач комп'ютерної лінгвістики.

Ключові слова: бритва Оккама, принцип, формальне поняття, природно-мовна конструкція, комп'ютерна лінгвістика, задача.

Abstract

The cases of possible use of the concept "Ockham's razor" for the analysis and synthesis of natural-language constructions are considered. It is proposed to formalize the concept of Ockham's razor for the purpose of exploit it as the criterion for solving a certain class of the tasks of computer linguistics.

Keywords: Ockham's razor, principle, formal concept, natural-language construction, computer linguistics, task.

Вступ

Найбільш освічені люди з пошту князів Коріатовичів, що засновували на Поділлі поселення Вінниця у середині XIV століття [1], мабуть, чули про свого сучасника – англійського філософа-схоласта, францисканця Вільяма з селища Оккама та читали його праці з логіки та фізики Аристотеля. Власне з часів останнього був відомий «принцип достатньої ґрунтовності», який сьогодні відомий під назвою «Бритва Оккама» (БО) завдяки грі слів у латинському оригіналі «*Entia non sunt multiplicanda praeter necessitatem*» – зголити також означає знайти істину, відокремити хибне припущення [2].

Актуальність цього методологічного принципу підтверджується прискіпливою увагою до нього протягом всього цього часу не тільки філософів і богословів, але й біологів, юристів, логіків та математиків. Зокрема становлено зв'язок поняття БО з фундаментальними математичними поняттями складності за Байесом та Колмогоровим [3], а також повноти за Геделем. Мають відношення до нього машини Тюрінга та Поста [4, 5]. Разом з алгоритмічною складністю бритва Оккама мігрувала в комп'ютерні науки, зокрема в теорії інформації стверджуються, що найбільш коротке повідомлення є найбільш надійним та сталим в процесах кодування-декодування [6]. У роботах [7, 8] розглядається тісний зв'язок між синтаксисом та семантикою теорії з огляду на їхню складність, проте недостатньо розкрито застосування принципу БО в моделюванні природно-мовної діяльності для задач комп'ютерної лінгвістики.

Мета дослідження полягає у визначенні напрямів застосування принципу бритви Оккама для розв'язання задач комп'ютерної лінгвістики.

Результати дослідження

В предметній області аналізу математичних теорій, зокрема формальних систем та граматик Хомського, застосування машин Тьюрінга і Поста та передаванні закодованої інформації у цілому йде мова про оброблення символічних послідовностей без урахування їхнього сенсу. Тобто суттєвим для аналізу та синтезу є лише синтаксис рядка слів з символів, а семантика не береться до уваги. Якщо такий підхід є ефективним, наприклад, для автоматизованого доведення теорем, то для оброблення природно-мовної інформації семантичні характеристики повідомлення (речення) мають ключове значення [9]. Застосування моделі асоціативної мережі мовних образів для розв'язання задач комп'ютерної лінгвістики [10] дозволило дослідити функціональні процеси природно-мовного обміну інформацією та відповідні критерії ефективності таких процесів. Розглянемо природу та приклади зазначених процесів з огляду на можливість використання принципу БО:

1. Опис певних подій або природних явищ від першої особи (без елементів діалогу) літературного або публіцистичного характеру. Зазвичай супроводжується засобами виразу почуттів та емоцій цієї особи до відтворюваних подій чи явищ. Зрозуміло, що скільки людей (авторів), стільки й особистих

думок, вражень, асоціацій тощо, так само як і у читачів відповідного тексту. Якщо одному читачу сподобається лаконічний стиль викладення, то іншому більш до душі буде відчуті спорідненість своїх емоцій та почуттів до авторських. У цьому випадку не варто стверджувати про доцільність використання принципу БО як формального критерія природно-мовного матеріалу.

2. Діалог певних осіб, наприклад, персонажів літературного твору. Як у житті, так і в літературі учасники діалогу обов'язково мають мотив, який намагаються досягнути шляхом перемовин. Процес діалогу зазвичай супроводжується набором емоцій його учасників, які викликаються обставинами, що супроводжують проміжні та кінцеві результати досягнення власного мотиву. Якщо розглядати чергову репліку певного учасника діалогу як ланцюг вербальних образів, кожний з яких пов'язаний зі сторонніми вербальними образами, то бритва Оккама дозволить швидше досягти цілі. Причина цього досить проста – кожна зайва ланка ланцюгу потенційно викликає у співбесідника непотрібні асоціації, особливо, якщо слова відповідної репліки мають емоційне забарвлення.

Можливо, саме тому людство завжди цінило лаконізм висловлювань як засіб досягнення мети за найкоротший час. Маємо вражаючі приклади «згортки» величезного обсягу відомої інформації як фактичного так і емоційного характеру в геніальних літературних творах. Так, назва «По кому дзвонить дзвін» вміщує весь зміст однойменного роману Є. Хемінгуей, а коротка репліка Майстра «Вільний!» до Понтія Пілата завершує неймовірний за гамою почуттів та глибиною матеріалу сюжет роману М. Булгакова «Майстер і Маргарита».

3. Ділова конструктивна мова за визначенням потребує мінімуму слів для донесення певної інформації. Прикладами є написи на табличках, накази, інструкції, технічні та наукові тексти тощо. Діловий стиль природної мови максимально наближений до формальних штучних мов, у т.ч. мов програмування. Надзвичайну вагу грає синтаксис природно-мовної конструкції, а бритва Оккама у даному випадку є засобом боротьби з багатозначністю слів та словосполучень. Ті ж самі анекдоти про програмістів дуже яскраво характеризують випадки невірної прийняття рішення за директивною вказівкою виключно тому, що за формальними ознаками остання включала в себе декілька можливих інтерпретацій.

Іншими прикладами такого роду висловлювань є відоме (рос.) «казнить нельзя помиловать», де місце коми є вирішальним для засудженого, або легенда про «Платонівську людину». Коли учні попросили Платона дати визначення людині, філософ сказав: «Людина є тварина на двох ногах, без пір'я». Почувши це, Діоген упіймав півня, обскував его, приніс в Академію та проголосив: «Ось Платонівська людина!» Після чого Платон додав до свого визначення: «Та з плоскими нігтями».

4. Випадок оперування інформацією у межах одного речення, наприклад, коли задається питання до відомого речення. З точки зору лінгвістики підмет та присудок виступають як остов будь-якого повного речення, тобто несуть у собі найбільше інформації, проте лаконізм відповіді може бути втіленим в одне єдине слово речення, що є вичерпною відповіддю на поставлене питання.

Критеріальну основу та формальну реалізацію такого випадку (без використання назви «Бритва Оккама», але за сутністю цього принципу) було докладно досліджено на основі запропонованої в [11] формальної теорії та моделі комутативної напівгрупи образних конструкцій. В [10] показано зв'язок між кількістю інформації, що містить природно-мовна конструкція з введеним на основі моделі асоціативної мережі мовних образів поняттям одиниці сенсу.

Висновки

Проведено дослідження виявило особливості використання принципу бритви Оккама для розв'язання задач комп'ютерної лінгвістики у чотирьох випадках, зокрема розглянуто 1) опис певних подій або природних явищ від першої особи; 2) діалог певних осіб; 3) ділова (інструктивна, конструктивна) мова та 4) випадок оперування інформацією у межах одного речення. Для трьох останніх випадків показано доцільність застосування принципу БО як критерія синтезу природно-мовних конструкцій, а для випадків 2) та 3) потребують розробки відповідні формальні моделі з метою реалізації такого критерію.

Подальшого дослідження також потребує визначення доцільності застосування принципу бритви Оккама для аналізу та синтезу текстів гумористичного спрямування. Оскільки, за переконанням автора, розуміння гумору формальною системою є частковим критерієм порівняння штучного інтелекту з природним на зразок тесту Тюрінга для діалогу, результати дослідження за цим напрямом мають високу актуальність.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Отамановський В. Д. Вінниця в XIV–XVII століттях: Історичне дослідження. – Вінниця, Континент-Прім, 1993. – 462 с.
2. Бритва Оккама [електронний ресурс] / Матеріал з Вікіпедії – вільної енциклопедії. – Режим доступу: https://uk.wikipedia.org/wiki/Бритва_Оккама.
3. Paul M. B. Vitányi and Ming Li; IEEE Transactions on Information Theory, Volume 46, Issue 2, Mar 2000 Page(s):446–464, "Minimum Description Length Induction, Bayesianism and Kolmogorov Complexity."
4. Chris S. Wallace and David L. Dowe; Computer Journal, Volume 42, Issue 4, Sep 1999 Page(s):270–283, "Minimum Message Length and Kolmogorov Complexity."
5. Scott Needham and David L. Dowe (2001):" Message Length as an Effective Ockham's Razor in Decision Tree Induction." Proc. 8th International Workshop on Artificial Intelligence and Statistics (AI+STATS 2001), Key West, Florida, U.S.A., Jan. 2001 Page(s): 253–260.
6. Nannen, Volker. "A short introduction to Model Selection, Kolmogorov Complexity and Minimum Description Length" (PDF). Archived (PDF) from the original on 2 June 2010. Retrieved 3 July 2010.
7. Alan Baker (2010) [2004]. "Simplicity". Stanford Encyclopedia of Philosophy. California: Stanford University. ISSN 1095-5054.
8. Quine, W V O (1961). "Two dogmas of empiricism". From a logical point of view. Cambridge: Harvard University Press. pp. 20–46. ISBN 0-674-32351-3.
9. Oleg V. Bisikalo, Roman N. Kvetny, Petro M. Povidayko, Yuriy A. Bunyak and others. Information Processing in Business, Security and Multimedia. Algorithms / Edited by Arkadiusz Liber. – Wrocław : Oficyna Wydawnicza Politechniki Wrocławskiej, 2013. – 129 p. – ISBN 978-83-7493-808-2.
10. Bisikalo O. Solving problems on base of concepts formalization of language image and figurative meaning of the natural-language constructs [Електронний ресурс] / Oleg Bisikalo, Slawomir Cieszczyk, Gulbahar Yussupova // 16th Conference on Optical Fibers and Their Applications. International Society for Optics and Photonics. – December 18, 2015. – Pp. 98161U-98161U-14.
11. Bisikalo O. Formalization of semantic network of image constructions in electronic content [Електронний ресурс] / O. Bisikalo, I. Kravchuk // Cornell University Library (Computer Science, Computation and Language), arXiv: 1201.1192v1. – January 2012. – 4 p. – Режим доступу: <http://arxiv.org/abs/1201.1192v1>.

Бісікало Олег Володимирович – д-р техн. наук, професор, декан факультету КСА, Вінницький національний технічний університет, м. Вінниця

Bisikalo Oleg V. – Dr.Sc. (Eng.), Professor, Dean of the Faculty for Computer Systems and Automatic, Vinnytsia National Technical University, Vinnytsia