

УДК 510.635:004.891(045)

doi:10.20998/2413-4295.2017.23.13

ЗАСТОСУВАННЯ ТЕМПОРАЛЬНОЇ ЛОГІКИ ПРИ ПОБУДОВІ ФОРМАЛЬНИХ МОДЕЛЕЙ ЕЛЕКТРОННИХ ТЕКСТІВ

А. І. ВАВИЛЕНКОВА

кафедра комп'ютеризованих систем управління, Національний авіаційний університет, Київ, УКРАЇНА
e-mail: vavilenkova@gmail.com

АНОТАЦІЯ Змістовний аналіз електронних текстових документів передбачає виявлення та встановлення логічних зв'язків між формальними моделями, якими описується зміст окремих частин – речень природної мови. Тому у статті проаналізовано засоби опису впорядкованості подій у часі, що базуються на темпоральній логіці, та запропоновано формальні правила та шаблони для автоматичного визначення часових рамок електронного текстового документу за рахунок відстеження ланцюжків тем і рем, тобто послідовності розгортання думки.

Ключові слова: логіко-лінгвістична модель; природна мова; темпоральна логіка; зміст; текст.

APPLICATION OF TEMPORAL LOGIC FOR CONSTRUCTION OF FORMAL MODELS OF ELECTRONIC TEXTS

A. VAVILENKOVA

Department of Computerized Control Systems, National Aviation University, Kyiv, UKRAINE

ABSTRACT The content analysis of electronic text documents involves identifying and establishing of logical links between formal models that describe the content of separate parts - the natural language sentences. Therefore, the purpose of the article is to develop formal rules and templates for the automatic determination of the time frame of an electronic text document by tracking themes and rhemes, that is, the sequence of unfolding (deployment) of thought. The study analyzes the existing means of setting the time frame in text documents based on temporal logic: first, to analyze the means of unfolding of thought in a text document, then to form the formal rules and templates to identify the thematic progress in electronic text documents. It is provided a geometric interpretation of three variants for template construction: for sequential unfolding of thought in the text, for one topic, which is repeated in each sentence of the text, and for the split topic. Unfolding thoughts in texts can take place in different ways, so the proposed rules and templates can be combined. If the constructed array of logic and linguistic models for the text fragment coincide with one of the types of the indicated templates, then it will be possible to automatically determine the nature of the unfolding of thought in this fragment. The article demonstrates an example of constructing one of the proposed templates for a text fragment. To Formulated formal conditions and developed templates are the basis for establishing content links in electronic text documents in general.

Keywords: logic and linguistic model; natural language; temporal logic; content; text.

Вступ

Змістовний аналіз електронних текстових документів передбачає виявлення та встановлення логічних зв'язків між формальними моделями, якими описується зміст окремих частин – речень природної мови [1]. Виявлення певних текстових прототипів, тобто зразків текстової побудови, є однією з операцій в типологічних дослідженнях текстів [2].

Структурний принцип організації довільного тексту ґрунтується на вживанні розділових знаків і пов'язаний зі змістовним принципом організації тексту, оскільки структурно значущі частини є одночасно і логічно значимими [3]. У зв'язку з цим при змістовному аналізі тексту необхідно враховувати ряд проблем, що виникають при визначенні предикативності, перерозподілу функціональної та інформативної значимості по мірі розгортання думки.

На сьогоднішній день для опису впорядкованості подій у часі використовується темпоральна логіка, тобто логічна система, що

дозволяє формалізувати твердження, істинність яких змінюється з часом, не вводячи явно поняття часу [4]. Приклад темпорального аналізу природної мови за Г. Райхенбахом можна відобразити у вигляді рис. 1 [5].

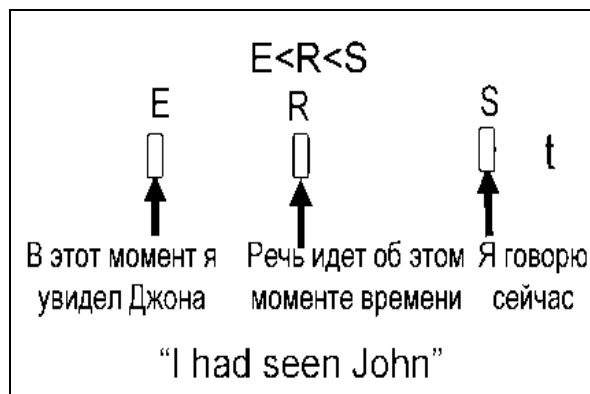


Рис. 1 – Приклад темпорального аналізу природної мови за Г. Райхенбахом

Тут R – точка референції, S – момент мови, E – час події або стану.

Темпоральну логіку можна також вважати одним із варіантів модальної логіки (логічної системи, в якій присутні модальні оператори), де оператор Fp означає "колись у майбутньому p виконається обов'язково", а Gp – "завжди в майбутньому буде виконуватися p ".

Наприклад, речення "Довільне відправлене повідомлення буде отримане" за допомогою модальних операторів буде представлено наступним чином:

$$G[\text{відправлене}(m) \Rightarrow F \text{отримане}(m)],$$

де m – означає "повідомлення".

Ще одним засобом простеження послідовності подій у текстових документах є логіка дерев обчислень, що базується на моделі Кріпке [6], де кожна вершина дерева відповідає за певний стан моделі, при цьому кожен наступний стан досягається лише із стану, з'єднаного з ним дугою (рис. 2) [7].

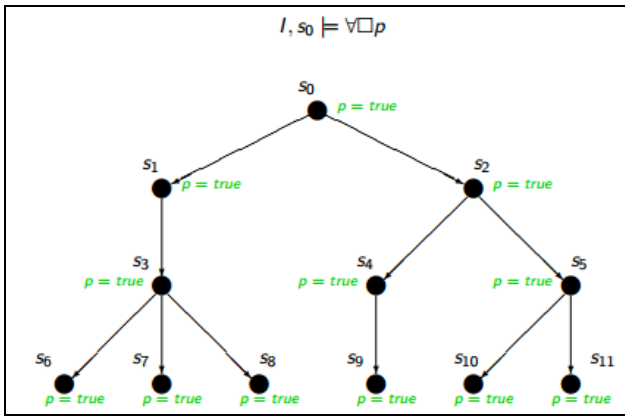


Рис. 2 – Приклад використання логіки дерев обчислень для опису послідовності подій

Проте для опису впорядкованості подій електронного тексту засобів темпоральної логіки не достатньо, так як існування певного тексту передбачає певну форму його організації, де розгортання думки відбувається від відомого до нового шляхом використання різноманітних засобів членування текстів. Саме на розробку формальних засобів виявлення впорядкованості подій в електронних текстових документах і спрямовані подальші дослідження.

Мета роботи

Метою роботи є розробка формальних правил та шаблонів для автоматичного визначення часових рамок електронного текстового документу за рахунок відстеження ланцюжків тем і рем, тобто послідовності розгортання думки. Для досягнення поставленої мети передбачається вирішення таких завдань:

- дослідити засоби встановлення часових рамок у текстових документах;
- дослідити формальні можливості опису послідовності дій в електронних текстових документах;
- проаналізувати засоби розгортання думки в текстових документах;
- сформулювати формальні правила та шаблони для виявлення тематичних прогресій в електронних текстових документах;
- за сформульованими правилами побудувати один із запропонованих шаблонів на основі використання логіко-лінгвістичних моделей речень текстового фрагменту.

Викладення основного матеріалу

Для відображення у структурі наукового тексту концептуального часу повинні узгоджуватися два принципи: статичний (коли описуються деякі часові істини, правила та закони) та динамічний (коли описується певний експеримент, об'єкт або дія). Для кожного з принципів використовується певний набір мовних засобів. Рух часових форм забезпечує зв'язність тексту. При створенні темпоральної структури наукового тексту ключову роль відіграє дієслово або відношення, що являється основою формування логіко-лінгвістичних моделей речень природної мови. Ядром функціонально-семантичної категорії темпоральності є мікросистема часових форм [8].

При цьому кожен текстовий фрагмент будується на основі певного типу тематичного зв'язку [9], комбінування яких забезпечує цілісність сприйняття тексту в цілому. За Ф. Данешом розрізняють п'ять типів абстрактних моделей, які лежать в основі побудови текстів [10]. На їх основі можна вивести логічні схеми (шаблони) формування змістовних зв'язків між реченнями природної мови.

Нехай текстовий фрагмент t представлено у вигляді масиву логіко-лінгвістичних моделей речень $t = \{L^1, L^2, \dots, L^S, \dots, L^{S_{N(t)}}\}$, кожна з яких має вигляд [11]:

$$L^S = \bigwedge_{z_{\rho} \in Z_{\rho}^S(x_{\rho}, g_{\rho}, y_{\rho}, q_{\rho}, h_{\rho})} \bigwedge_{r_{\rho} \in R_{\rho}^S(x_{\rho}, g_{\rho}, y_{\rho}, q_{\rho}, z_{\rho}, h_{\rho})} p_{\rho}(x_{\rho}, g_{\rho}, y_{\rho}, q_{\rho}, z_{\rho}, r_{\rho}, h_{\rho}).$$

Якщо для фрагменту тексту характерне **послідовне розгортання** думки, тобто нова інформація з попереднього речення стає темою для наступного речення, тобто об'єкти або предмети відношень кожного попереднього речення можуть виступати суб'єктами, об'єктами та предметами відношень кожного наступного. Виявлення можливих комбінацій передбачає створення шаблонів масивів логіко-лінгвістичних моделей типу:

$$t = \{L^1, L^2, \dots, L^S, \dots, L^{S_{N(t)}}\} = \{p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) \&$$

$$p_2(z_1, g_2, y_2, q_2, z_2, r_2, h_2) \& \dots \&$$

$$p_\delta(x_\delta, g_\delta, y_{\delta-1}, q_\delta, z_\delta, r_\delta, h_\delta) \& \dots \&$$

$$p_{N(t)}(z_{N(t)-1}, g_{N(t)}, y_{N(t)}, q_{N(t)}, z_{N(t)}, r_{N(t)}, h_{N(t)})\}.$$

Число створених шаблонів дорівнює кількості комбінацій із можливих компонентів логіко-лінгвістичних моделей, що підлягають заміні, у кожному наступному реченні, тобто $x_{\delta+1}, y_{\delta+1}$ або $z_{\delta+1}$ із кожного наступного на y_δ або z_δ попереднього. При цьому граматична характеристика часу для кожного із відношень логіко-лінгвістичних моделей однакова, тобто $h(p_1) = h(p_2) = \dots = h(p_\delta) = \dots = h(p_{N(t)})$.

Геометрична інтерпретація послідовного розгортання думки у тексті зображена на рис.3.

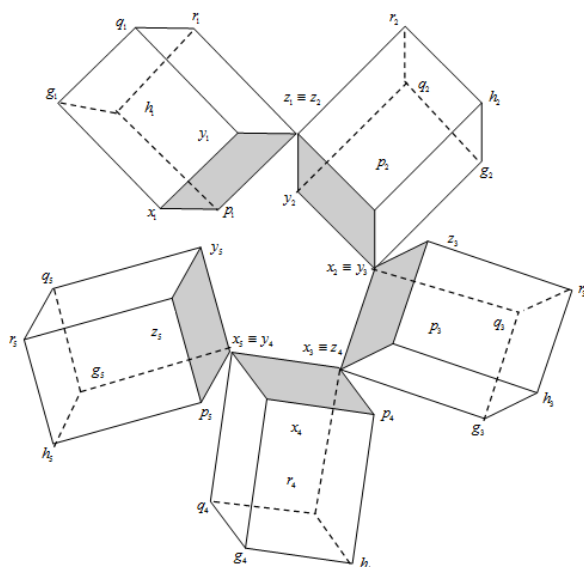


Рис. 3 – Геометрична інтерпретація послідовного розгортання думки у тексті

Якщо для фрагменту тексту характерна наявність **однієї теми, яка повторюється у кожному реченні тексту**, то для нього буде побудовано шаблон з масиву логіко-лінгвістичних моделей такого типу:

$$t = \{L^{S_1}, L^{S_2}, \dots, L^{S_\delta}, \dots, L^{S_{N(t)}}\} =$$

$$\{p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) \&$$

$$p_2(x_1, g_2, y_2, q_2, z_2, r_2, h_2) \& \dots \&$$

$$\dots \dots \dots$$

$$p_\delta(x_1, g_\delta, y_\delta, q_\delta, z_\delta, r_\delta, h_\delta) \& \dots \&$$

$$\dots \dots \dots$$

$$p_{N(t)}(x_1, g_{N(t)}, y_{N(t)}, q_{N(t)}, z_{N(t)}, r_{N(t)}, h_{N(t)})\}.$$

Компонентою, що буде присутня у логіко-лінгвістичних моделях всіх наступних речень

фрагменту може бути x_1, y_1 або z_1 із збереженням їх характеристик g_1, q_1 або r_1 відповідно.

Геометричну інтерпретацію для одного із можливих варіантів вживання теми, яка повторюється у кожному реченні тексту, показано на рис. 4.

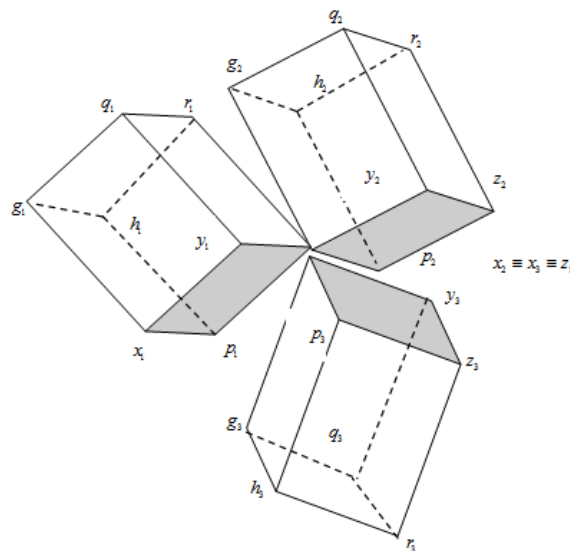


Рис. 4 – Геометрична інтерпретація варіанту вживання теми, яка повторюється в кожному реченні тексту

На відміну від першого шаблону, граматична характеристика часу для першого та всіх наступних речень не однакова, тобто

$$h(p_2) = \dots = h(p_\delta) = \dots = h(p_{N(t)}) = h(p_1) - 1 \text{ або}$$

$$h(p_2) = \dots = h(p_\delta) = \dots = h(p_{N(t)}) = h(p_1) + 1.$$

Якщо у фрагменті тексту **тема розщеплена**, тобто в декількох наступних реченнях присутні початкові точки, від яких думка розвивається далі, то шаблон з масиву логіко-лінгвістичних моделей буде мати наступний тип:

$$t = \{L^{S_1}, L^{S_2}, \dots, L^{S_\delta}, \dots, L^{S_{N(t)}}\} =$$

$$\{p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) \&$$

$$p_2(x_1, g_2, y_2, q_2, z_2, r_2, h_2) \& \dots \&$$

$$\dots \dots \dots$$

$$p_\delta(y_{\delta-1}, g_\delta, y_\delta, q_\delta, z_\delta, r_\delta, h_\delta) \& \dots \&$$

$$\dots \dots \dots$$

$$p_{N(t)}(y_{\delta-1}, g_{N(t)}, y_{N(t)}, q_{N(t)}, z_{N(t)}, r_{N(t)}, h_{N(t)})\}.$$

Компонентами логіко-лінгвістичних моделей, що можуть повторюватися в реченнях, від яких розвивається думка x_1, y_1 або z_1 , при цьому суб'єктом кожного наступного речення може бути відповідно $y_{\delta-1}$ або $z_{\delta-1}$. При цьому граматична характеристика часу співпадає з граматичною характеристикою часу відношення, від якого розвивається думка, тобто $h(p_{\delta-1}) = \dots = h(p_\delta) = \dots = h(p_{N(t)})$.

Геометричну інтерпретацію для одного із можливих варіантів вживання теми, що розщеплюється, показано на рис. 5.

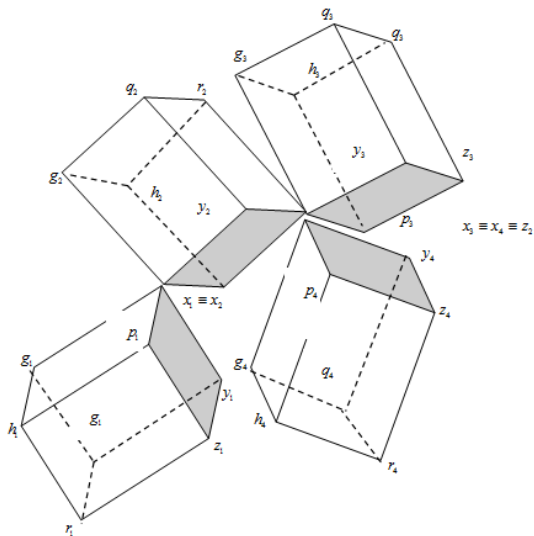


Рис. 5 – Геометрична інтерпретація варіанту вживання розщепленої теми

Розгортання думки в текстах може відбуватися різними шляхами, тому запропоновані правила та шаблони можуть комбінуватися. Відповідно, шаблони можна надбудувати.

Обговорення результатів

Нехай необхідно формально описати впорядкованість подій у часі для заданого текстового фрагменту: "Публіцистичний стиль займає проміжне місце між науковим і художнім стилями. Основним його призначенням є вплив на слухача або читача. У текстах публіцистичного стилю часто загальноживані слова набувають нового звучання".

По-перше, необхідно побудувати логіко-лінгвістичну модель текстового фрагменту, шляхом побудови логіко-лінгвістичних моделей речень природної мови.

Логіко-лінгвістична модель речень для цього фрагменту буде мати вигляд:

$$t = \{L^{S_1}, L^{S_2}, L^{S_3}\} =$$

$$\{L^{S_1} = p_1(x_1, g_1, y_1, q_1, z_1, r_{11}, h_1) \& p_1(x_1, g_1, y_1, q_1, z_1, r_{12}, h_1),$$

$$L^{S_2} = p_2(x_2, g_{21}, y_2, q_2, z_2, r_2, 0) \vee p_2(x_2, g_{22}, y_2, q_2, z_2, r_2, 0),$$

$$L^{S_3} = p_3(x_3, g_3, y_3, q_3, z_3, r_3, h_3)\}.$$

$$t = \{L^{S_1}, L^{S_2}, L^{S_3}\} =$$

$$\{L^{S_1} = \text{займає(стиль, публіцистичний, місце, проміжне, стилем науковим, між)} \&$$

$$\text{займає(стиль, публіцистичний, місце, проміжне, стилем, художнім, між)}.$$

$$L^{S_2} = \epsilon(\text{вплив, слухача, призначенням, основним, стиль, публіцистичний, 0}) \vee$$

$\epsilon(\text{вплив, читача, призначенням, основним, стиль, публіцистичний, 0})$.

$L^{S_3} = \text{набувають(слова, загальноживані, звучання, нового, текстах, публіцистичного \& стилю, часто)}$.

Другим кроком є синтез логіко-лінгвістичних моделей речень, внаслідок чого елементи логіко-лінгвістичних моделей другого і третього речення заданого текстового фрагменту будуть замінені на відповідні елементи логіко-лінгвістичної моделі першого речення $z_2 \equiv x_1, r_2 \equiv g_1, r_3 \equiv x_1 \& g_1$, тобто:

$$t^{\gamma} = \{L^{S_1}, L^{S_2}, L^{S_3}\} =$$

$$\{L^{S_1} = p_1(x_1, g_1, y_1, q_1, z_1, r_{11}, h_1) \& p_1(x_1, g_1, y_1, q_1, z_1, r_{12}, h_1),$$

$$L^{S_2} = p_2(x_2, g_{21}, y_2, q_2, x_1, g_1, 0) \vee p_2(x_2, g_{22}, y_2, q_2, x_1, g_1, 0),$$

$$L^{S_3} = p_3(x_3, g_3, y_3, q_3, z_3, x_1 \& g_1, h_3)\}.$$

Наступним етапом є визначення граматичних характеристик часу для відношень кожного з трьох речень: $h(p_1) = h(p_2) = h(p_3) = \text{теперішній}$.

Заключний етап – це звірка отриманого формального опису з запропонованими шаблонами. Рівність граматичних характеристик відношень, а також заміни, здійснені під час синтезу, дають змогу віднести заданий текстовий фрагмент до першого типу шаблону, що відповідає послідовному розгортанню думки.

Висновки

У статті наведено опис засобів, що використовуються сьогодні для простеження часу виконання подій в текстових документах та в основі яких лежить темпоральна логіка. Враховуючи те, що при створенні темпоральної структури наукового тексту ключову роль відіграють дієслова або відношення логіко-лінгвістичних моделей, автором запропоновано шаблони масивів таких моделей. Якщо побудований для текстового фрагменту масив логіко-лінгвістичних моделей буде співпадати з одним із типів наведених шалонів, то можна буде автоматично встановити характер розвитку думки у даному фрагменті. Сформульовані формальні умови та розроблені шаблони є базою для встановлення змістовних зв'язків в електронних текстових документах в цілому.

Список літератури

1. **Graneheim, U.H.** Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness / **U.H. Graneheim, B. Lundman** // *Nurse Education Today*. – 2004. – Vol. 24, №2. – P. 105 – 112.
2. **Чернявская, З. Е.** Лингвистика текста: поликодовость, интертекстуальность, интердискурсивность / **Чернявская З. Е.** – М.: Книжный дом «Либроком», 2009. – 248 с.
3. **Jurafsky, D.** Natural Language Processing / **D. Jurafsky, C. Manning**. – Stanford University, 2012. – Available at: <https://www.coursera.org/course/nlp/> – 15.05.2014.
4. **Lampert, L.** Specifying Systems: The TLA+ Language and

- Tools for Hardware and Software Engineers / **Lampert L.** – Boston: Addison-Wesley Longman Publishing Co., 2002. – 364 p.
- Карпов, Ю. Г.** Математическая логика и теория алгоритмов / **Карпов Ю. Г.** – Санкт-Петербургский политехнический университет. – Режим доступа: <http://dcn.icc.spbstu.ru>. – 24.12.2016.
 - Laroussinie, F.** On Model Checking Durational Kripke Structures / **F. Laroussinie, N. Markey, Ph. Schnoebelen** // *Foundations of Software Science and Computation Structures*. – 2002. – P. 264 – 279. – doi: 10.1007/3-540-45931-6_19.
 - Захаров, В. А.** Основы математической логики и логического программирования. – Режим доступа: <http://mk.cs.msu.ru/images/8/80/LectLog18.pdf>.
 - Geeraerts, D.** Cognitive linguistics: basic readings research / **Geeraerts D., Dirven R., John R.** – Berlin-New York: Publ. mouton de cruyter, 2006. – 486 p.
 - Evans, V.** Lexical concepts, cognitive models and meaning-construction / **V. Evans** // *Journal of Cognitive semiotics*. – 2006. – P. 73–107.
 - Филиппов, К. А.** Лингвистика текста: курс лекций / **Филиппов К. А.** – Спб.: Издательство С. – Петербургского университета, 2008. – 336 с.
 - Вавиленкова, А. И.** Основные принципы синтеза логико-лингвистических моделей / **А. И. Вавиленкова** // *Кибернетика и системный анализ*. – 2015. – Т. 51, № 5. – С. 176 – 185.
 - measures to achieve trustworthiness, *Nurse Education Today*, 2004, Vol.24, №2, 105 – 112.
 - Chernyavskaya, Z. E.** Lingvistika teksta: polikodovost', intertekstual'nost', interdiskursivnost', M., Knizhnuy dom "Librokom", 2009, 248p.
 - Jurafsky, D., Manning, C.** Natural Language Processing, Stanford University, 2012, Available at: <https://www.coursera.org/course/nlp/> – 15.05.2014.
 - Lampert, L.** Specifying Systems: The TLA+ Language and Tools for Hardware and Software Engineers, Boston: Addison-Wesley Longman Publishing Co., 2002, 364 p.
 - Karpov, U. G.** Matematicheskaya logika i teoriya algoritmov, Sankt-Peterburgskiy politexnicheskii universitet, Available at: <http://dcn.icc.spbstu.ru>.
 - Laroussinie, F., Markey, N., Schnoebelen, Ph.** On Model Checking Durational Kripke Structures, *Foundations of Software Science and Computation Structures*, 2002, 264 – 279, doi: 10.1007/3-540-45931-6_19.
 - Zaxarov, V.A.** Osnovu matematicheskoy logiki o logicheskogo programmirovaniya, Available at: <http://mk.cs.msu.ru/images/8/80/LectLog18.pdf>.
 - Geeraerts, D., Dirven, R., John, R.** Cognitive linguistics: basic readings research, Berlin-New York: Publ. mouton de cruyter, 2006. – 486 p.
 - Evans, V.** Lexical concepts, cognitive models and meaning-construction, *Journal of Cognitive semiotics*, 2006, 73–107.
 - Filippov, K. A.** Lingvistika teksta:kurs lekcziy, Izdatel'stvo Sankt-Peterburgskogo universiteta, 2008, 336 p.
 - Vavilenkova A. I** Osnovnue prinzcipu sinteza logiko-lingvisticheskix modeley, *Kibernetika i sistemnyy analiz*, 2015, 51, № 5, P. 176 – 185.

Bibliography

- Graneheim, U.H., Lundman, B.** Qualitative content analysis in nursing research: concepts, procedures and

Відомості про авторів (About authors)

Вавиленкова Анастасія Ігорівна – кандидат технічних наук, доцент, Національний авіаційний університет, доцент кафедри комп'ютеризованих систем управління; м. Київ, Україна; e-mail: vavilenkovaa@gmail.com.

Anastasia Vavilenkova – Candidate of Technical Sciences (Ph. D.), Docent, Associate Professor, Department of Computerized Control Systems, National Aviation University, Kyiv, Ukraine; e-mail: vavilenkovaa@gmail.com.

Будь ласка, посилайтеся на цю статтю наступним чином:

Вавиленкова, А. І. Застосування темпоральної логіки при побудові формальних моделей електронних текстів / **А. І. Вавиленкова** // *Вісник НТУ «ХПІ»*, Серія: Нові рішення в сучасних технологіях. – Харків: НТУ «ХПІ». – 2017. – № 23 (1245). – С. 84-88. – doi:10.20998/2413-4295.2017.23.13.

Please cite this article as:

Vavilenkova, A. Application of temporal logic for construction of formal models of electronic texts. *Bulletin of NTU "KhPI". Series: New solutions in modern technologies*. – Kharkiv: NTU "KhPI", 2017, 23 (1245), 84–88, doi:10.20998/2413-4295.2017.23.13.

Пожалуйста, ссылайтесь на эту статью следующим образом:

Вавиленкова А. И. Применение темпоральной логики при построении формальных моделей электронных текстов / **А. И. Вавиленкова** // *Вестник НТУ «ХПИ»*, Серія: Новые решения в современных технологиях. – Харьков: НТУ «ХПИ». – 2017. – № 23 (1245). – С. 84-88. – doi:10.20998/2413-4295.2017.23.13.

АННОТАЦИЯ Содержательный анализ электронных текстовых документов предусматривает выявление и установление логических связей между формальными моделями, с помощью которых описывается содержание отдельных частей – предложений естественного языка. Поэтому в статье проанализированы способы описания упорядочивания событий во времени, которые основаны на темпоральной логике, и предложены формальные правила и шаблоны для автоматического определения временных рамок электронного текстового документа за счет отслеживания цепочек тем и рем, то есть последовательности развития мысли.

Ключевые слова: логико-лингвистическая модель; естественный язык; темпоральная логика; содержание; текст.

Надійшла (received) 31.05.2017