

М.К. МОРОХОВЕЦЬ

Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: marina.morokhovets@gmail.com.

Н.М. ЩОГОЛЕВА

Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: natashch2904@gmail.com.

ЗАСТОСУВАННЯ ТА РОЗРОБЛЕННЯ ЗАСОБІВ ЛІНГВІСТИЧНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ АНАЛІЗУ ТА ПЕРЕТВОРЕННЯ ТЕКСТІВ ВИМОГ ДО ПРОГРАМНИХ СИСТЕМ

Анотація. Проведено експериментальні дослідження текстів вимог до програмних систем за допомогою частотного аналізу. За результатами експериментів виділено «мову таймерів» — сукупність речень природної мови, що мають певну структуру та описують дії з таймерами. Розроблено засоби перетворення речень мови таймерів у позначені структуровані вирази проміжної мови, яку можна використовувати як допоміжний засіб у ланцюжку кроків від тексту вимог до системи, поданого природною мовою, до формальної моделі цієї системи.

Ключові слова: частотний аналіз, мовний процесор, формалізація спеціалізованих текстів, мова таймерів.

ВСТУП

У роботах [1, 2] описано методику розроблення функціональних специфікацій систем та пристроїв, а також систему оброблення натуральномоовних специфікацій (СОНС), що призначена для аналізу текстів вимог до програмних систем та перетворення їх на тексти, подані формальною мовою.

У цій статті описано нові засоби, розроблення яких було заплановано для розвитку мовного процесора СОНС (див. [2]), а саме: засоби формалізації нових видів речень, поповнення вихідних мов мовного процесора засобами опису структури речень відповідних вхідних мов.

Перехід від опису вимог до програмної системи, поданих природною мовою, до формальної моделі цієї системи є одним з етапів розроблення програмних систем. Кожна система має складову часу, адже з часом пов'язаний такий важливий аспект, як поведінка системи.

Формальна модель системи будується на основі тексту вимог і має бути узгоджена з цими вимогами. Перш ніж обирати засоби формалізації, зокрема, часових вимог, природно з'ясувати, які саме вимоги у тексті пов'язані з часом.

У межах дослідження текстів вимог до програмних систем щодо складової часу опрацьовано низку текстів різного обсягу. Інструментом дослідження текстів на початковому етапі обрано частотний аналіз, у результаті якого утворюється частотний список слів тексту. Фахівці використовують частотний аналіз у різних галузях для дослідження текстів, поданих природною мовою. Наприклад, частотний аналіз літературних текстів застосовано на підготовчому етапі під час роботи над повним словником Шевченкової мови [3, С. 354–355]; як показали експерименти із системою статистичного оброблення професійних текстів [4], частотний аналіз дає змогу визначати тематику науково-технічних та галузевих текстів.

У цій роботі частотний аналіз застосовано для пошуку слів (словоформ), пов'язаних з часом, що трапляються в аналізованому тексті вимог найчастіше. У результаті проведених експериментів виявилось, що таким словом є «timer»

(таймер, тобто засіб, що відраховує інтервали часу). На наступних етапах роботи з аналізованими текстами спочатку було виділено ті речення, що містять це слово та його словоформи, а потім вивчено структуру виділених речень та побудовано мову таймерів. Зрештою розроблено модифікацію мовного процесора СОНС, а також проміжну мову позначених фраз. У результаті здобуто засіб перетворення фраз мови таймерів у позначені структуровані фрази цієї проміжної мови. Надалі проміжну мову можна застосовувати як вхідну для перетворювачів фраз мови таймерів у вирази тієї мови, яку обирають для формалізації.

ЗАСОБИ ЧАСТОТНОГО АНАЛІЗУ ПРИРОДНОМОВНИХ ТЕКСТІВ

Для здійснення частотного аналізу у цій роботі використано програмну систему LIS (Linguistic System), розроблену на базі мовних процесорів DUET і FEST [5] та призначену для статистичного оброблення текстів, поданих природною флективною мовою. Система LIS дає змогу будувати частотні списки словоформ (фраз) вхідного тексту, що стосуються певної галузі (головно науково-технічних текстів) [4, 6]. Спосіб формування вихідного частотного списку залежить від наповнення інформаційної бази системи та значень вхідних параметрів.

Інформаційна база LIS складається зі спеціальних таблиць та словників різних видів.

Словники та таблиці будують засобами LIS за формальними специфікаціями морфології та лексики мови вхідних текстів. Специфікації подають спеціальними формальними мовами Lmorf та Lduet [5, 7]. Мова Lmorf — це метамова для опису морфологічної інформації; мова Lduet — декларативна метамова, що є засобом формального опису (специфікації) лексики вхідної мови разом з граматичною інформацією, необхідною для оброблення текстів, поданих цією вхідною мовою. Обмеження області застосування скорочує словник і запобігає синонімії.

Функціональні можливості LIS є такими:

- побудова морфологічних таблиць за специфікаціями морфології вхідної мови, поданими мовою Lmorf; специфікації будує фахівець-морфолог;
- побудова словників за специфікацією лексики мовою Lduet;
- морфологічний аналіз вхідного тексту та побудова списку словоформ та словосполучень, виведення побудованого списку у вихідний файл;
- побудова частотного списку словоформ вхідного тексту на основі результатів морфологічного аналізу цього тексту;
- побудова частотного списку фраз вхідного тексту на основі результатів морфологічного аналізу цього тексту;
- генерація специфікацій лексем мовою Lduet за частотним списком та генерація словоформ цих лексем.

Специфікація лексики — це подання слів, що не змінюються, та основ слів, що змінюються, разом з їхніми граматичними характеристиками.

Нині інформаційна база LIS складається з морфологічних таблиць кількох різних мов, словників лексики різних мов. За допомогою LIS можна будувати різні словники, а побудовані словники можна поповнювати [8].

Для налаштування LIS на роботу з текстами певної природної мови потрібно:

- побудувати морфологічний опис цієї мови засобами Lmorf;
- побудувати за цим описом морфологічні таблиці за допомогою LIS;
- здійснити специфікацію лексики цієї мови засобами Lduet.

Налаштовану на роботу з текстами певної природної мови систему можна поповнювати (засобами LIS) різноманітними словниками (як-от: словниками

термінів з різних галузей знань, словником загальноновживаних слів) та використовувати у сеансі роботи з LIS той чи інший словник з наявних у системі.

Засоби LIS дають можливість будувати частотні списки словоформ з вхідного тексту у різних режимах — залежно від словника, що використовується у сеансі роботи з LIS, або незалежно від словника, наприклад:

— виділяти та включати у частотний список, що будується, слова, знайдені у словнику (побудова частотного списку так званих «відомих» слів);

— не виділяти й не включати у частотний список слова, знайдені у словнику (побудова частотного списку так званих «невдомих» слів);

— виділяти та включати у частотний список усі слова вхідного тексту.

ВИКОРИСТАННЯ РЕЗУЛЬТАТІВ ЧАСТОТНОГО АНАЛІЗУ ТЕКСТІВ ВИМОГ

За допомогою системи LIS здійснено частотний аналіз кількох текстів описів програмних систем. Розглянемо хід роботи на прикладі оброблення фрагмента знайденого у відкритому доступі тексту [9]. Для зручності викладення дамо цьому робочому фрагменту назву TS. Обсяг тексту TS становить 49 сторінок файлу Word у doc-форматі. Перший крок роботи з TS — застосування програми частотного аналізу. В експерименті використано систему LIS зі словником загальноновживаних слів, який було застосовано для аналізу технічних наукових текстів. Режимом використання LIS був пошук «невдомих» слів (такі слова характеризують специфіку тексту, що досліджується). У результаті отримано частотний список слів (точніше, словоформ) вхідного тексту.

Програма частотного аналізу формує результат у вигляді списку словоформ разом із супутньою інформацією. Кожен рядок списку містить такі відомості: порядковий номер рядка, кількість входжень слова та його модифікацій у вхідний текст (тобто у текст, який аналізують), відсоток входжень слова та його модифікацій відносно загальної кількості слів у тексті, номер рядка у вхідному файлі, в якому слово чи його модифікація трапилися вперше, слово та його модифікації. Наведемо початковий фрагмент результату частотного аналізу тексту TS, тобто побудованого за TS частотного списку. Він обмежений першим входженням словоформи, що стосується часу, та має такий вигляд:

1)	527	3.016%	24	messag (-e, -es)
2)	398	2.277%	24	BS (-s)
3)	310	1.774%	3	req
4)	304	1.740%	28	MSC
5)	222	1.270%	7	shall
6)	217	1.242%	28	circuit (-0, -s)
7)	150	0.858%	4	procedur (-e, -es)
8)	130	0.744%	28	resourc (-e, -es)
9)	117	0.669%	28	HANDOVER
10)	97	0.555%	24	information
11)	96	0.549%	28	radio
12)	89	0.509%	6	REQUEST
13)	82	0.469%	15	handover
14)	81	0.463%	28	interfac (-e)
15)	78	0.446%	7	connection (-0, -s)
16)	76	0.435%	48	call (-s, -0)
17)	74	0.423%	48	channel (-0, -s)
18)	73	0.418%	28	ol (-d)
19)	72	0.412%	164	CIRCUIT
20)	64	0.366%	28	indicat (-ed, -e, -es)

21)	63	0.360%	79	block (-0, -ed)
22)	62	0.355%	6	ASSIGNMENT
23)	60	0.343%	40	sent
24)	57	0.326%	28	allocat (-e, -ed, -es)
25)	57	0.326%	67	valu (-e, -es)
26)	55	0.315%	176	ACKNOWLEDG (-e)
27)	54	0.309%	63	FAILUR (-e)
28)	54	0.309%	91	COMMAN (-d)
29)	52	0.298%	40	terrestrial
30)	52	0.298%	48	caus (-e, -es)
31)	45	0.257%	10	request (-0, -ed, -s)
32)	45	0.257%	32	element (-0, -s)
33)	44	0.252%	52	sending
34)	43	0.246%	460	VB (-s)
35)	42	0.240%	115	SCCP
36)	41	0.235%	28	GPP
37)	41	0.235%	48	sen (-d)
38)	40	0.229%	32	pool (-s, -0)
39)	39	0.223%	28	timer (-0, -s).

Першим словом, що стосується часу, у наведеному частотному списку є «timer» (таймер). Як видно з наведеного списку, це слово та його словоформи зустрічаються у тексті TS 39 разів. У інших аналізованих під час досліджень текстах слово «timer» має високу частоту входжень, що свідчить про вагомість відповідного поняття у цих текстах. З огляду на це було досліджено контексти цього слова у текстах.

Подальший аналіз тексту TS та деяких інших природномовних текстів, що містять опис вимог до проєктованих програмних систем, дав змогу виділити «мову таймерів», тобто сукупність речень, що описують дії з таймерами. Це здійснено за допомогою мовного процесора СОНС, налаштованого на пошук у тексті та відбір речень, що містять слово «timer» (таймер) та його словоформи. Після цього дослідник має можливість проаналізувати структуру речень цієї мови, аби подати цю мову у формі вхідної граматики мовного процесора СОНС. Далі описано розроблені засоби аналізу та оброблення речень цієї мови.

Аналіз контекстів слів, що стосуються часу, є кроком на шляху переходу від опису, поданого природною мовою, до формальної моделі проєктованого об'єкту.

ЗАСОБИ СОНС ДЛЯ ОБРОБЛЕННЯ РЕЧЕНЬ МОВИ ТАЙМЕРІВ

Під мовою таймерів розуміємо сукупність речень природної мови, що мають певну структуру та описують дії з таймерами. Не описуючи докладно синтаксис мови таймерів, подамо загальну структуру її речень. Речення мови таймерів є прості та складені. Структура простих речень є такою:

<дія> timer <ім'я таймера>,
<ім'я суб'єкта> <дієслівна група> timer <ім'я таймера>,
<дія> timers <ім'я таймера> and <ім'я таймера>,

де *дія* подається дієсловом *start, clear, stop, cancel, ім'я таймера* — це просте або складене ім'я; *просте ім'я* — ідентифікатор, *складене ім'я* має вигляд *ідентифікатор:ідентифікатор* або *ідентифікатор-ідентифікатор*, *ім'я суб'єкта* — це просте ім'я (ідентифікатор) або складене ім'я (послідовність ідентифікаторів), перед яким може стояти артикль, *дієслівна група* — це фраза «shall clear» або «shall stop».

Структура складених речень є такою:

<дія> timer <ім'я таймера> <умова>,
<дія> timer <ім'я таймера> <мета>,
<ім'я суб'єкта> <дієслівна група> timer <ім'я таймера> and <подальша дія>,
<подія> <ім'я суб'єкта> <дієслівна група> timer <ім'я таймера>.

Здійснено налаштування СОНС на синтаксичний аналіз речень мови таймерів. На базі СОНС розроблено засоби перетворення речень мови таймерів на структуровані вирази, у яких складові вхідних речень мають позначки. За допомогою створених засобів здійснюється синтаксичний аналіз вхідного речення. Якщо він завершується успішно, то речення перетворюється на вираз, структура якого може відрізнитися від будови цього речення, а фрази вхідного речення позначаються. Наведемо приклади перетворень речень мови таймерів за допомогою розроблених засобів.

Речення

```
Clear timer ХНТ3.
```

перетворюється у такий спосіб:

```
type: clear-t-sent  
Clear timer  
    timer-name  
ХНТ3.
```

Дія з таймером — `clear timer` — визначає тип вхідного речення та результату його перетворення; опис типу (`type: clear-t-sent`) вміщується на початку виразу, що є результатом перетворення. Ім'я таймера (`ХНТ3`) отримує позначку `timer-name`.

Речення

```
The XC shall clear timer ХНТ5.
```

перетворюється у такий спосіб:

```
type: sh-clear-t-sent  
    subject-name  
XC  
    shall-action  
shall clear timer  
    timer-name  
ХНТ5.
```

Перший рядок описує тип речення, другий рядок містить позначку імені суб'єкта дії з таймером (`subject-name`), після позначки слідує ім'я суб'єкта (`XC`), далі вказано позначку дії з таймером (`shall-action`), за нею слідує сама дія, два останні рядки містять позначку імені таймера (`timer-name`) та ім'я таймера (`ХНТ5`).

Речення

```
Clear timer ХНТ3, if running.
```

перетворюється у такий спосіб:

```
type: if-clear-t-sent  
    condition  
timer  
    timer-name  
ХНТ3  
    timer-state  
is running.
```

```
action
Clear timer
timer-name
ХНТЗ.
```

Щойно розглянуте речення є умовним. Такі речення перетворюються на вирази усталеної структури [1], в яких спочатку (під позначкою `condition`) подається умова (ця частина виразу завершується крапкою), а за нею (після позначки `action`) слідує дія. Ім'я таймера (`ХНТЗ`) позначене (`timer-name`) в розділі умови й у розділі дії. Умова у цьому випадку стосується стану таймера (`is running`). Стан таймера має позначку (`timer-state`). Особливістю вхідного речення є те, що в умові (`if running`) не вказано ані тип об'єкта, ані його ім'я. А у виразі-результаті розділ умови містить і тип об'єкта, й ім'я. Під час перетворення речень такого типу, крім додавання позначок до фраз вхідного речення, здійснюється також поповнення його частин та модифікація його структури.

У подібний спосіб перетворюються речення

```
Stop timer МНТ if it's running.
Stop timer В:МНТ if it is still running.
```

Речення

```
Stop timers SHТ and SLТ.
```

перетворюється у такий спосіб:

```
type: stop-ts-sent
Stop timer
timer-name
SHТ.
Stop timer
timer-name
SLТ.
```

Тут опис типу виразу (`type: stop-ts-sent`) вказує на те, що після нього слідує два підвирази однакової структури.

Речення

```
Start timer МНТ1 to guard against the C: Target Channel Ready
message from the SU-SF.
```

перетворюється у такий спосіб:

```
type: start-t-sent
Start timer
timer-name
МНТ1
```

```
purpose
to guard against
message-name
C: Target Channel Ready
from-addr-name
SU-SF.
```

Тут, крім дії з таймером й імені таймера, позначено мету дії та складові вхідного речення, що стосуються опису цієї мети.

Речення

```
On receipt of a HANDOVER SUCCEEDED from the M1, the old B1 shall
stop timer T8.
```

перетворюється у такий спосіб:

```
type: on-rec-shstop-t-sent
```

```
On receipt of
  receipt-object
HANDOVER SUCCEEDED
  from-addr-name
M1
  receipt-subject
old B1
  shall-action
shall stop timer
  timer-name
T8.
```

Тип виразу у першому рядку (`type: on-rec-shstop-t-sent`) вказує на те, що дія з таймером (зупинення таймера) залежить від події отримання деякого об'єкта, який позначено у виразі за допомогою фрази `receipt-object`. Також позначками виділено відправника та адресата (об'єкта).

Перетворення речення, поданого природною мовою, на вираз з позначками, дає змогу подати вхідне речення мовою, яка має риси формальної та природної. У виразі-перекладі, як видно з наведених прикладів, явно описується його тип та структура. Опис структури та типу виразів-перекладів формалізовано. За допомогою позначок створюється «структурний каркас» виразу-перекладу, а «наповнення» цього каркасу здійснюється змістовими складовими вхідного речення, поданими природною мовою. Зазначимо, що мова виразів-перекладів не є природною, адже речення природної мови не містить опису своєї структури; ця мова не є формальною, бо вирази-переклади містять фрази природною мовою. Назвемо мову виразів-перекладів проміжною. Використання проміжної мови спрощує переклад з природної мови на формальну мову. До того ж за наявності проміжної мови можна відкласти остаточний вибір формальної мови, якою буде подана модель проєктованого об'єкта, а у разі заміни формальної мови подання об'єкта іншою (що може відбутися під час реінжинірингу) полегшується перехід до іншої мови.

ПІДСУМКИ ТА ВИСНОВКИ

За допомогою системи LIS здійснено частотний аналіз низки текстів вимог до програмних систем з метою пошуку словоформ, що стосуються часу та трапляються у текстах часто. Проведені експерименти показали високу частоту входжень у тексти словоформ слова «timer». Мовний процесор СОНС було налаштовано на пошук у вхідному тексті речень, що містять слово «timer» та його словоформи, й побудову вихідного тексту, що складається з усіх таких речень. Подальший аналіз добутих речень зі словоформами слова «timer» дав змогу виділити типи таких речень, визначити структуру речень кожного типу та скласти «мову таймерів», тобто сукупність речень природної мови, що описують дії з таймерами. Ця мова була подана у вигляді вхідної граматики мовного процесора СОНС.

Для кожного типу речень мови таймерів визначено форму виразу, у який мовний процесор має перетворювати речення даного типу. Для забезпечення потрібних перетворень розроблено низку семантичних дій, що доповнили сукупність семантичних дій мовного процесора СОНС. Розроблені нові семантичні дії дають змогу, зокрема, замінити у межах речення займенник іменем об'єкта. Також складено послідовності семантичних дій, якими декоровано правила побудованої вхідної граматики. Розроблено проміжну мову позначених

фраз, якою подаються речення вхідної мови у результаті оброблення мовним процесором. Ця мова названа проміжною, бо розглядається як допоміжний засіб у ланцюжку кроків від тексту вимог до системи, поданого природною мовою, до формальної моделі цієї системи. Допоміжними для подальшої роботи елементами проміжної мови є описи типів речень, позначки структурних частин речення, позначки імен таймерів та суб'єктів дій з ними, станів таймерів та інших змістових частин вхідного речення, виділених на етапі синтаксичного аналізу.

Зазначимо, що позначання фраз речень є інструментом оброблення текстів, поданих природною мовою, що використовується у різних галузях для різних потреб. Наприклад, у роботі [10], у галузі інженерії вимог (requirement engineering), позначання (анотування) використано для того, щоб виділяти змістові складові речень юридичних текстів, а також зазначено, що таке позначання важливе для переходу від юридичних текстів до формальних специфікацій або моделей. У роботах [1, 11] позначання застосовано для виділення структурних складових текстів та речень з метою підготовки їх до подальшого машинного оброблення.

Зауважимо також, що обидві системи, LIS та СОНС, використані у цій роботі, розроблено на базі Розширеної Системи Програмування ТЕРЕМ [12], інструменту для розроблення мовних процесорів.

Проведені з використанням частотного аналізу експериментальні дослідження текстів вимог до програмних систем стосовно складової часу дали змогу виділити «мову таймерів», мовний процесор СОНС налаштовано на перетворення нових видів речень, а саме речень мови таймерів, у позначені структуровані вирази проміжної мови.

Використання проміжної мови дає можливість:

- здійснити підготовчу роботу до перетворення речень вхідної мови на формальні вирази;
- відтермінувати обрання формальної мови для подання речень вхідної мови, якщо є різні думки щодо того, яку цільову формальну мову слід обрати;
- зменшити зусилля під час реінжинірингу готової системи у тому разі, коли виникає потреба використати іншу формальну мову.

У роботі показано застосування засобів лінгвістичного програмного забезпечення (системи LIS та СОНС) для аналізу та перетворення спеціалізованих природномовних текстів. Система LIS будує частотний список слів тексту, за допомогою якого можна, зокрема, визначити тематику тексту, виявити у тексті поняття з певної галузі знань. Засоби СОНС дають змогу описувати, зокрема, так звані галузеві мови та будувати перетворювачі речень цих мов на позначені вирази, у яких позначками виділено змістові складові речень.

Проведені експерименти показали, що СОНС та LIS виявилися корисними підручними інструментами для дослідження галузевих текстів та здійснення підготовчих кроків до побудови формальних моделей за цими текстами. Зокрема, подання речень галузевого тексту мовою позначених фраз є частковою формалізацією цього тексту. Накопичення більшого досвіду аналізу та оброблення галузевих текстів дасть змогу розширити можливості СОНС для формалізації текстів, поданих природною мовою.

Автори планують продовжити дослідження текстів вимог до програмних систем щодо складової часу, а також розглянути застосування та розширення засобів СОНС для різних видів оброблення природномовних текстів інших профілів.

СПИСОК ЛІТЕРАТУРИ

1. Годлевский А.Б., Мищенко Н.М., Мороховец М.К., Фелижанко О.Д., Щёголева Н.Н. Методика разработки и средства формализации функциональных спецификаций систем и устройств. *Кибернетика и системный анализ*. 2016. Т. 52, № 5. С. 25–42.
2. Мищенко Н.М., Мороховец М.К., Фелижанко О.Д., Штелик Е.В., Щёголева Н.Н. Новые функциональные возможности системы обработки естественных языковых спецификаций и среда её функционирования. *Кибернетика и системный анализ*. 2018. Т. 54, № 6. С. 37–46.
3. Іван Огієнко (Митрополит Іларіон). Тарас Шевченко. Тимошик М.С (Ред.). Київ: Наша культура і наука, 2002. 440 с.
4. Мищенко Н.М., Федюрко В.В., Щёголева Н.М. Мобильная система статистической обработки профессиональных текстов в целях определения их тематики. *Проблемы программирования*. 1999. № 2. С. 11–18.
5. Мищенко Н.М., Фелижанко О.Д., Щёголева Н.Н. Универсальная система программ обработки тематических текстов. *УСiМ*. 2012. № 6. С. 35–42.
6. Мищенко Н.М. Про інтеграцію средств пословной обработки текстов на основе формальных спецификаций лексики. *Тезисы докладов международной научной конф. «Искусственный интеллект»* (11–16 сентября 2000, Кацивели, Крым, Украина). Кацивели, 2000. *Искусственный интеллект*. 2000. № 3. С. 114–116.
7. Міщенко Н.М. Система програм морфологічного аналізу науково-технічних текстів. *Зб. «Наукові записки». Матеріали п'ятої Міжнародної науково-практичної конференції «Мови і світ: дослідження та викладання»*. (24–25 березня 2011, Кіровоград, Україна). Кіровоград, 2011. *Серія: Філологічні науки*. 2011. Вип. 95 (2). С. 538–542.
8. Міщенко Н.М., Феліжанко О.Д., Щоголева Н.М. Засоби розширення граматичного словника лексики за частотним списком невідомих слів. *Зб. «Наукові записки». Матеріали п'ятої Міжнародної науково-практичної конференції «Мови і світ: дослідження та викладання»*. (24–25 березня 2011, Кіровоград, Україна). Кіровоград, 2011. *Серія: Філологічні науки*. 2011. Вип. 95 (2). С. 543–547.
9. Digital cellular telecommunications system (Phase 2+); Mobile-services Switching Centre - Base Station System (MSC - BSS) interface; Layer 3 specification (3GPP TS 08.08 version 8.7.0 Release 1999). URL: https://www.etsi.org/deliver/etsi_ts/100500_100599/100590/08.07.00_60/ts_100590v080700p.pdf.
10. Sleimi A., Sannier N., Sabetzadeh M., Briand L., Ceci M., Dann J. An automated framework for the extraction of semantic legal metadata from legal texts. *Empirical Software Engineering*. 2021. Vol. 26, Iss. 3. 43. <https://doi.org/10.1007/s10664-020-09933-5>.
11. Лети́чевський О.О., Мороховець М.К., Щоголева Н.М. Деякі засоби обробки електронних текстових документів. *Control Systems and Computers*. 2021. № 4. С. 13–18. <https://doi.org/10.15407/csc.2021.04.013>.
12. Мищенко Н.М., Щёголева Н.Н. О проектировании языковых процессоров на ПЭВМ. *Кибернетика и системный анализ*. 1993. Т. 29, № 2. С. 110–117.

М.К. Morokhovets, N.M. Shchogoleva

APPLICATION AND DEVELOPMENT OF LINGUISTIC SOFTWARE TOOLS FOR ANALYSING AND TRANSFORMING THE TEXTS OF REQUIREMENTS FOR SOFTWARE SYSTEMS

Abstract. Experimental analysis of the texts of requirements for software systems by means of frequency analysis is carried out. The “timer language,” i.e., a set of natural-language sentences of definite structure that describe actions with timers, is distinguished according to the results of the experiments. Means for transforming sentences of timer language into marked structured expressions of intermediate language are developed. The intermediate language proposed can be used as an auxiliary mean in the chain of steps from the natural-language text of requirements for a system to the formal model of this system.

Keywords: frequency analysis, language processor, formalization of specialised texts, timer language.

Надійшла до редакції 12.05.2022