

Адам Пшепюрковський

Використання програми
POLIQRAR
для пошуку в корпусі
(на прикладі корпусу
IPi PAN)

Переклала Ореста Тимчишин

Кафедра формальної лінгвістики
Варшавський університет
Варшава 2008

Оригінал:
Adam Przepiórkowski
Rozdział 4. Przeszukiwanie korpusu
в
Korpus IPI PAN. Wersja wstępna
Instytut Podstaw Informatyki PAN
Warszawa 2004
ISBN 83-910948-8-X

Copyright © 2004 Adam Przepiórkowski. Усі права застережені
Copyright © 2008 Ореста Тимчишин (український переклад)

Цей документ є вільним. Ви можете розповсюджувати його та/або вносити зміни відповідно до умов Загальної Публічної Ліцензії GNU FDL у тому вигляді, у якому вона була опублікована Фондацією Вільного Програмного Забезпечення; або 2-ї версії Ліцензії, або (на ваш розсуд) будь-якої більш пізньої версії.

Aknowlegments

A translation is performed within scope of the International Visegrad Fund master's scholarship program — <http://www.visegradfund.org/>.

I want to thank:

- to the author of the original book Dr. Adam Przepiórkowski for his agreement to publish the translation under the terms of the GNU Free Documentation Licence and for the support
- to Prof. Janusz Stanisław Bień supervising of my work
- to the team of Department of Formal Linguistics of the University of Warsaw and its Head Prof. Magdalena Maria Danielewiczowa and to other people from DoFL for a nice reception and informal, but very stimulant working atmosphere
- to Dr. Natalia Kotsyba for numerous advices and fruitful discussions which often went beyond scopes of translation

Подяки

Переклад виконано в рамках магістерської стипендіальної програми Міжнародного Вишеградського Фонду (International Visegrad Fund — <http://www.visegradfund.org/>).

Хочу висловити подяку:

- автору Адаму Пшепіорковському за підтримку і згоду опублікувати переклад на засадах GNU Free Documentation License
- професору Варшавського університету Янушу Станіславу Беню за координацію роботи
- колективу кафедри формальної лінгвістики Варшавського університету та її завідувачу професору Магдалені Марії Данелевічовій за теплий прийом і неформальну, але дуже стимулюючу робочу атмосферу
- Наталі Коцибі за численні поради і плідні дискусії, які часто виходили за рамки власне перекладу

Розділ 4. Пошук у корпусі

Для пошуку в Корпусі IPI PAN і відображення конкордансів служить програма Poliqar¹, створена в рамках цього проекту Зигмунтом Криніцкім і Даніелем Янусом, під керівництвом автора². Poliqar має амбіції сати універсальним інструментом для пошуку в корпусах: тагсет (множина тегів), використаний в цьому корпусі, не є вбудований у програму, а просто задається через зовнішній конфігураційний файл. Внутрішній формат кодування знаків — це універсальний формат UTF-8. Інших перешкод для використання цієї програми для різних корпусів, в тому числі, іншомовних, немає.

Програма Poliqar існує у трьох версіях:

- графічна версія, описана в п. 2.2, для наступних операційних систем: Windows 2000, Windows XP (а також, можливо, для інших Windows-систем, але не була в них протестована), і GNU/Linux;
- текстова версія, описана в п. 2.3, для операційної системи GNU/Linux;
- internet-версія, описана в п. 2.1, дозволяє користуватися Корпусом IPI PAN через Internet, використовуючи довільний браузер, наприклад, Mozilla, Internet Explorer, Opera чи Links.

Спільною для цих трьох версій є багата мова запитів, описана в п. 1.

1 Мова запитів

Мова запитів у програмі Poliqar базується на системі, використаній у програмі Corpus Query Processor (CQP) створеній у Штутгартському Університеті (Christ, 1994), але містить ряд розширень і покращень. Цей розділ описує систему запитів програми Poliqar і ілюструє її багатьма прикладами.

1.1 Запити про сегменти

У найпростішому випадку запит це послідовність сегментів, які треба знайти, наприклад:

- (1) przyszedł czas
- (2) przyszedł em rano

У запиті (2) виступають три сегменти, що відповідають двом словам: *przyszedłem* і *rano*. У випадку простих запитів про форми Poliqar пробує виявити випадки, де слово складається з кількох сегментів, і знайти відповідні послідовності сегментів, отже правильні результати дадуть, також запити:

- (3) przyszedłem rano
- (4) długom szedł

У другому випадку буде знайдено всі випадки трьохсегментних послідовностей

d	l	u	g	o
---	---	---	---	---

m

s	z	e	d	ł
---	---	---	---	---

, що будуть інтерпретуватися як прислівник, аглютинант і псевдодієприкметник, як і в двосегментових послідовностях

d	l	u	g	o	m
---	---	---	---	---	---

s	z	e	d	ł
---	---	---	---	---

,

¹ POLyinterpretation Indexing Query and Retrieval Processor.

²На початкових етапах роботи над проектуванням та впровадженням також брав участь Матеуш Пшепорковскі.

інтерпритованих як іменникова форма у давальному відмінку і псевдодієприкметник.

За замовчуванням розрізняється регістр (*kasztowość*³) букв, а отже, наведені нижче запити дадуть різні результати:

(5) przyszedł

(6) Przyszedł

Щоб знайти всі входження форми *przyszedł*, ігноруючи величину окремих букв, треба вжити прапорець /i. Наведені нижче запити дадуть однакові результати, що будуть містити всі відповіді на два вище наведені запити.

(7) przyszedł/i

(8) Przyszedł/i

У графічній і текстовій версіях програми чутливість до величини букв можна також задати глобально — для цілого запиту, або ряду запитів (див. п. 2.2 і 2.3).

У запитах про сегменти можуть виступати стандартні регулярні вирази, в яких використовуються наступні спеціальні символи: ?, *, +, ., ,, |, {, }, [,], (,), а також натуральні числа, записані арабськими цифрами, наприклад, 0 чи 21, при чому специфікації сегментів, що містять регулярні вирази, беруться в лапки ". Але формальний опис регулярних виразів не входить в рамки цієї публікації, тому обмежимося тут кількома прикладами, які мають дозволити користувачеві швидко засвоїти правила запису і значення таких виразів.

(9) "Ala|Ela"

знак | означає альтернативу двох виразів, тому цей запит може бути використаний для знаходження всіх входжень сегментів *Ala* або *Ela*,

(10) "[AE]la"

квадратні дужки означають альтернативу знаків, отже, цей запит можна вживати для знаходження сегментів, в яких першим знаком є *A* або *E*, а потім іде послідовність *la*, тобто цей запит є аналогічним до попереднього,

(11) "beza?"

знак питання означає необов'язковість знаку (у даному випадку *a*) або взятого в круглі дужки виразу, що знаходиться безпосередньо перед ?, отже в результаті задання даного запиту будуть знайдені сегменти *bez* і *beza*,

(12) "bez."

крапка означає довільний знак, результатом такого запиту будуть сегменти *beza*, *bezy*, *bezq* і т. д., але не *bez* чи *bezami*,

(13) "bez.?"

bez, *beza*, *bezy*, *bezq* і т. д., але не *bezami*,

(14) ".z.z."

п'ятизнакові сегменти, в яких 2-ий і 4-ий знак — це знак *z* (наприклад *czczq* і *rzezi*),

(15) ".z.z..?"

сегменти, що складаються з п'яти або шести знаків, в яких 2-ий і 4-ий знаки — це *z*, наприклад *czczq*, *rzezi* і *szczyt*,

³Цей термін, відповідник англійському *case sensitivity*, був запропонований у працях (Bień, 1991, 2001).

- (16) "a*by"
зірочка означає будь-яку кількість входжень знаку або виразу безпосередньо перед нею, отже, цей запит слугуватиме для знаходження сегментів, що складаються з довільної кількості букв *a*, після яких виступає послідовність *by*, наприклад, *by* (нуль входжень *a*), *aby*, *aaaaby* і т. д.,
- (17) "Ala.*"
сегменти, що починаються на *Ala*, наприклад, *Ala* і *Alabama*,
- (18) "ala.*"/i
сегменти, що починаються на *ala*, *Ala*, *aLa*, *ALA* і т. д., наприклад, *Ala*, *alabaster* і *ALABAMA*,
- (19) ".*al+"
плюс виконує функцію, подібну до зірочки і означає довільну, *більшу від нуля* кількість входжень знаку, або виразу безпосередньо перед ним, результатом цього запиту будуть сегменти, що закінчуються на *al*, *all*, *alll* і т. д., але не на *a*, наприклад *dal*, *robal* і *Gall*,
- (20) "a{1,3}b.*"/i
конструкція типу {*n,m*} означає від *n* до *m* входжень знаку або виразу безпосередньо перед нею, цей запит допоможе знайти сегменти, що починаються послідовністю від 1 до 3 букв *a* або *A*, після яких іде буква *b* або *B*, а потім довільна послідовність знаків, наприклад, *Aby*, *aaaby*, *absolutnie*, *ABBA*,
- (21) ".*(1a){3,}.*"
конструкція типу {*n*,} означає мінімум *n* входжень знаку, або взятого у круглі дужки виразу безпосередньо перед нею, цей запит можна використати для знаходження сегментів, в яких послідовність *la* виступає принаймні 3 рази підряд, наприклад, *tralalala*, *sialalala*,
- (22) "[bc̣cḍdfghjkḷlṃṇńpṛṣsṭwẓẓ]{4,}[ạąę̣ę̣ịọụy]" /i
сегменти, що складаються з як мінімум 4 приголосних і однієї голосної, наприклад, *źdźbła* і *Chrzczę* — вираз [bc̣cḍdfghjkḷlṃṇńpṛṣsṭwẓẓ]{4,} означає принаймні чотириразове повторення знаку з [bc̣cḍdfghjkḷlṃṇńpṛṣsṭwẓẓ], тобто хоча б 4 входження приголосної (не обов'язково тієї самої),
- (23) "([bc̣cḍdfghjkḷlṃṇńpṛṣsṭwẓẓ]{3}[ạąę̣ę̣ịọụy]){2,}" /i
сегменти, що складаються з щонайменше дворазового повторення візрця CCCV, де C — це приголосна, а V — це голосна, наприклад, *wszystko*, *Zdmuchnąwszy* і *Szmajdziński* — конструкція типу {*n*} означає *n* входжень знаку, або взятого в круглі дужки виразу безпосередньо перед нею,
- (24) "(pod|na|za)jecha.*"
сегменти, що починаються з *podjecha*, *najecha* і *zajecha*, наприклад, *podjechał*, *zajechałszy*.

Наведені вище специфікації сегментів мають відповідати цілим сегментам — тому з обох сторін послідовності (1a){3,} в запиті (21) про сегменти, що містять послідовність *lalala* стоять послідовності *.**, що означають довільну послідовність знаків. Такого самого ефекту буде досягнуто при вживанні прапорця /*x*, що означає, що одна специфікація сегменту має відповідати принаймні частині даного сегменту, не обов'язково цілому сегменту:

- (25) "(1a){3,}" /*x*
сегменти, в яких послідовність *la* виступає принаймні 3 рази підряд, наприклад, *tralalala*, *sialalala*,
- (26) "(1a){3,}" /*ix*
сегменти, які містять послідовність *lalala*, *LaLaLa* і т. д., наприклад, *tralalala*, *SiaLaLaLa*.

1.2 Запити про початкові форми

Щоб знайти всі форми лексеми KORPUS, можна задати наступний запит:

(27) [base=korpus]

Атрибут `base` є одним з багатьох можливих атрибутів, які можуть виступити в запиті. Значенням цього атрибуту має бути специфікація початкової форми, отже, запит [base=pisać] можна задавати для пошуку форм типу *pisac*, *pisze*, *pisala*, *piszcie*, *pisanie*, *pisano*, *pisane* і т. д.

Іншим можливим атрибутом є `orth` — значення цього атрибуту окреслюють самі сегменти, отже наступні пари запитів дадуть однакові результати:

(28) a. przyszedł
b. [orth=przyszedł]

(29) a. Przyszedł/i
b. [orth=Przyszedł/i]

(30) a. przyszedł czas
b. [orth=przyszedł] [orth=czas]

Однак, не є однаковими запити:

(31) a. przyszedłem rano
b. [orth=przyszedłem] [orth=rano]

У першому випадку програма спробує розбити слово *przyszedłem* на два сегменти, *przyszedł* і *em*, тоді як значення атрибуту `orth` завжди інтерпретується як специфікація одного сегменту.

Значеннями атрибутів `base` і `orth` можуть бути регулярні вирази, подібно до випадку простих запитів про форми, описаних в п. 1.1, наприклад:

(32) [orth="bez.?" /i]
будуть знайдені сегменти *bez*, *Beza*, *bezy* і т. п., але не *bezm* чи *bezami*,

(33) [base="bez.?" /i]
будуть знайдені всі сегменти, початкова форма яких містить 3 або 4 букви і які починаються на *bez* (при довільній величині букв), наприклад, сегменти *bzem*, *bez*, *bezami* і т. п.

1.3 Запити вищого порядку

Запити про сегменти і про початкові форми сегментів можна поєднувати. Наприклад, щоб знайти всі входження сегменту *minę*, що інтерпретується як форма лексеми MINA (а не, наприклад, лексеми MIJAĆ), можна задати наступний запит:

(34) [orth=minę & base=mina]

Подібне значення має наступний запит про входження сегменту *minę*, які не інтерпретуються як форми лексеми MIJAĆ.

(35) [orth=minę & base!=mijać]

Умову про те, що початковою формою є не *mijać*, можна задати також використавши знак заперечення (знак оклику) перед назвою атрибуту, отже, наступний запит є рівнозначний попередньому:

(36) [orth=minę & !base=mijać]

Як і в логіці висловлень, запит з подвійним запереченням є тотожним запиту без заперечення, тому наступні запити про сегмент *nie* як форму особового займенника ON повністю рівноцінні:

- (37) [orth=nie & base=on]
 (38) [orth=nie & !base!=on]
 (39) [orth=nie & !!!base!=on]
 (40) [orth=nie & !!base=on]

В попередніх запитах оператор & виконує роль логічної кон'юнкції. Для позначення логічної альтернативи або вживається оператор |. Декілька прикладів вживань цього оператора:

- (41) [base=on | base=ja]
 всі форми займенників ON і JA, що відповідають запити [base="on|ja"],
 (42) [base=on | orth=mnie | orth=ciebie]
 всі форми займенника ON, а також сегменти MNIE і CIEBIE,
 (43) [orth=pora & !(base=por | base=pora)]
 сегмент *pora*, що не є ані формою лексеми POR, ані формою лексеми PORA.

Щоб краще зрозуміти різницю між операторами & і |, порівняймо два наступні запити:

- (44) [orth=minę & base=mina]
 (45) [orth=minę | base=mina]

В результаті задання першого запити будуть знайдені ті сегменти, які є одночасно (кон'юнкція) сегментом *minę* і формою лексеми MINA, а значить, тільки ті входження сегменту *minę*, які інтерпретуються як форми лексеми MINA. В результаті задання другого запити буде знайдено ті сегменти, які є або довільно інтерпретованими сегментами *minę*, або формою лексеми MINA (альтернатива), тобто всі входження сегменту *minę*, як і сегментів *mina*, *minu*, *minami* і т. д., інтерпретованих як форми лексеми MINA.

Специфікації позицій в корпусі взяті в квадратні дужки можуть містити довільну кількість умов типу **атрибут=значення** (наприклад, **orth=nie**), об'єднаних операторами **!**, **&** і **|**, як це показують вище наведені приклади. Також можливою є повна відсутність будь-яких умов — наступний запит може послужити для пошуку всіх сегментів у корпусі.⁴

- (46) []

Така тривіальна специфікація позиції у корпусі, що підходить до будь-якого сегменту, може допомогти, наприклад, у знаходженні двох форм, розділених двома довільними сегментами, наприклад:

- (47) [orth=się] [] [] [base=bać]

В результаті такого запити будуть знайдені послідовності, як, наприклад, *się nikogo nie bać*, *się Boga nie boicie* і т. п.

Для багатьох випадків цікавою була б можливість запити, наприклад, про форми, відстань між якими становить *максимум* п'ять позицій. Poliqgr дає можливість задання таких запитів, оскільки дозволяє формулювати регулярні вирази на рівні позиції корпусу. Наприклад, запит про форму лексеми ВАС, що виступає на дві, три або чотири позиції після форми SIE, може мати такий вигляд:

- (48) [orth=się] [] {2,4} [base=bać]

⁴Міг би, як би не той факт, що Poliqgr містить внутрішнє обмеження на кількість можливих результатів на запит.

В результаті цього запити будуть знайдені ті самі послідовності, що і в попередньому запиті, а також, наприклад, послідовність *się każdy następny Rywlin będzie bał*.

Звичайно, правильнішим запитом про різні входження форм так званого зворотнього дієслова ВАС SIE буде запит про *się* на певній відстані перед формою лексеми ВАС, але без знаку пунктуації між цими формами або безпосередньо після такої форми, можливо відділеної від форми ВАС особовим займенником:⁵

(49) [orth=się] [orth!="[.!?,:]"] {,5} [base=bać]
| [base=bać] [base=ón|ja|ty|my|wy]" ? [orth=się]

1.4 Запити про морфосинтаксичні теги

Запит (49) можна спростити, замінивши умову `orth!="[.!?,:]"` безпосереднім викликом “граматичного класу” `interp`:

(49') [orth=się] [pos!=interp] {,5} [base=bać]
| [base=bać] [base=ón|ja|ty|my|wy]" ? [orth=się]

Значеннями атрибуту `pos` (англ. *part of speech* ‘частина мови’) є скорочення назв граматичних класів. Наприклад, запит про послідовність двох іменникових форм, що починаються на *a* можна сформулювати наступним чином:

(50) [pos=subst & orth="a.*"] {2}

Як і у випадку специфікації текстових і початкових форм, специфікація граматичних класів може містити регулярні вирази. Наприклад, зважаючи на те, що особові займенники належать до класу третьоособових займенників `ppron3` і до класу нетретьоособових займенників `ppron12`, наступні запити можна використати для знаходження довільних форм будь-яких особових займенників:

(51) [pos=ppron12 | pos=ppron3]

(52) [pos="ppron12|ppron3"]

(53) [pos="ppron(12|3)"]

(54) [pos="ppron[123]+"]

(55) [pos="ppron.+"]

(56) [pos=ppron/x]

А запит (49) можна ще більше спростити до наступного запити:

(49'') [orth=się] [pos!=interp] {,5} [base=bać]
| [base=bać] [pos=ppron/x]" ? [orth=się]

В запитах можна окреслювати не тільки значення форм, що виступили в тексті (за допомогою атрибуту `orth`), початкових форм (за допомогою `base`) і граматичних класів (за допомогою `pos`), а також значення окремих граматичних категорій, таких як відмінок чи рід. Для цього служать наступні атрибути:

⁵Цей запит було розбито на два рядки з типографічних причин

атрибу́т	категорія	можливі значення
number	число	sg pl
case	відмінок	nom gen dat acc inst loc voc
gender	рід	m1 m2 m3 f n
person	особа	pri sec ter
degree	ступінь	pos comp sup
aspect	вид	imperf perf
negation	заперечення	aff neg
accentability	акцентування	akc nakc
post-prepositionality	поприйменниковість	npaep praep
accommodability	акомодативність	congr rec
agglutination	аглютинативність	agl nagl
vocalicity	вокальність	nwok wok

Отже, можна задавати, наприклад, наступні запити:

- (57) [number=sg]
буде знайдено всі форми в однині,
- (58) [pos=subst & number=sg]
буде знайдено всі іменникові форми в однині,
- (59) [pos=subst & gender!=f]
іменникові форми чоловічого або середнього роду,
- (60) [number=sg & case="nom|acc"& gender="m[123]"]
форми чоловічого роду у називному чи знахідному відмінку однини.

Замість повних назв атрибутів можна вживати їх скорочення з трьох букв:

атрибу́т	скорочення
number	nmb
case	cas
gender	gnd
person	per
degree	deg
aspect	asp
negation	neg
accommodability	acm
accentability	acn
post-prepositionality	ppr
agglutination	agg
vocalicity	vc1

Отже, запит(60) можна задати також наступним чином:

- (60') [nmb=sg & cas="nom|acc"& gnd="m[123]"]

Графічна і текстова версії дають можливість формулювання так званих аліасів, або альтернативних скорочень значення даного атрибуту, які потім можна вживати як звичайні значення атрибутів. У даній версії Корпусу ІРІ РАН означено наступні чотири аліаси:

аліас	оголошення
masc	m1 m2 m3
noun	subst depr ger xxs ppron12 ppron3
pron	ppron12 ppron3 siebie
verb	fin praet aglt bedzie inf imp3 impt pact ppas pcon pant ger winien

При такому означенні аліасів noun і masc два наступні запити є рівнозначними:

- (61) [pos=noun & gender=masc]
 (62) [pos="subst|depr|ger|xxs|ppron12|ppron3"
 & gender="m1|m2|m3"]

Про граматичні класи і категорії можна також задавати запити одночасно, використовуючи для цього атрибут `tag`. Наприклад, щоб знайти всі іменники середнього роду в називному відмінку однини, можна задати наступний запит:

- (63) [tag=subst:sg:nom:n]

Значення атрибуту `tag` має вигляд $k1:kat_1:kat_2:\dots:kat_n$, де $k1$ це назва граматичного класу, а kat_i це значення категорій, притаманних даному класу, в тій послідовності, в якій вони подані в наступній таблиці на стор. 11.

Як і у випадку інших атрибутів, специфікацію атрибуту `tag` можна задавати регулярним виразом:

- (60'') [tag=".*:sg:(nom|acc):m[123].*"]
 (60''') [tag="sg:(nom|acc):m[123]"/x]

1.5 Багатозначності і уоднозначнення

Однією з рис, які виділяють Корпус IPI PAN і програму Poliqarp з-поміж інших, є можливість репрезентації морфо-синтаксичних багатозначностей, згідно з постулатом, виголошеним в праці Oliva 2001. Існують конструкції, в яких неможливо уоднозначнити інтерпретацію даної форми, наприклад тому, що кілька можливих інтерпретацій зумовлює ідентичність значення цілого речення. Це ілюструє приклад, наведений нижче.

- (64) Pamiętam ją pijańą.
 (65) a. Pamiętam go pijańego.
 b. Pamiętam go pijańum.

У цьому прикладі не можна ствердити, що форма *pijańą* є у знахідному відмінку, як в *pijańego* в (65a), чи формою в орудному відмінку, як *pijańum* в (65). Тому в корпусі сегмент *pijańą* в (64) має отримати дві інтерпретації: в орудному і знахідному відмінках.

В такому випадку виникає питання: сегмент *pijańą* в (64) має бути знайдено в випадку задання запиту типу [case=acc], чи ні? Так, якщо брати до уваги той факт, що інтерпретація знахідного відмінку є правильною в цьому контексті. Ні, оскільки не можна з впевненістю сказати, що це форма знахідного відмінку.

Щоб не нав'язувати одну з цих інтерпретацій, Poliqarp дозволяє задавати обидва види запитів. Результатом запиту [case=acc] буде знаходження всіх сегментів, в яких принаймні однією із правильних інтерпретацій в даному контексті є інтерпретація в знахідному відмінку, отже, у випадку задання запиту (64) буде знайдено як сегмент *pijańą*, так і сегмент *ją*, але у випадку задання (65a) будуть знайдені обидві форми *go* і *pijańego*. Щоб знайти ті сегменти, які є в знахідному відмінку на 100 відсотків, тобто ті, для яких у даному контексті правильною є інтерпретація в знахідному відмінку, треба використати запит [case==acc]. Результатом останнього запиту в випадку (64) буде форма *ją*, але не форма *pijańą*, а у випадку задання (65a) буде заново знайдено форми *go* і *pijańego* — хоч ці форми мають також інтерпретацію родового відмінку, в контексті речення (65a) мають бути уоднозначненні до інтерпретації знахідного відмінку.

Завдяки цьому розрізненню можна, наприклад, шукати форми, які можуть бути у даному контексті інтерпретовані як форми у знахідному або родовому відмінку, отже, у правильно розміченому корпусі наступний запит може дати не пусті результати.

	число	відмінок	рід	особа	ступінь	вид	заперечення	акцентування	поприйменниковість	акомодативність	аглютинативність	вокальність
іменник	+	+	@									
депреціативний іменник	@	+	@									
кількісний числівник	@	+	+							+		
збірний числівник	@	+	@							+		
прикметник	+	+	+		+							
приприймен. прикметник												
післяприймен. прикметник												
прислівник					+							
нетретьоособовий займенник	@	+	+	@				+				
третьоособовий займенник	+	+	+	@				+	+			
займенник SIEBIE		+										
форма не мин. часу	+			+		@						
форма майб. часу ВУС	+			+		@						
аглютинант ВУС	+			+		@						+
псевдоімеслів	+		+			@					+	
наказовий спосіб	+			+		@						
безособова форма дієслова						@						
інфінітив						@						
дієприслівник активного стану						@						
дієприслівник пасивного стану						@						
віддієслівний іменник	+	+	@			@	+					
дієприкметник активного стану	+	+	+			@						
дієприкметник пасивного стану	+	+	+			@	+					
winiен	+		+			@						
предикатив												
прийменник		@										
сполучник												
kublik												
ciło obce nominalne	+	+	@									
ciąło obce luźne												
нерозпізнана форма												
інтерпункція												

(66) [case=acc & case=gen]

З іншого боку, наступний запит, що є запитом про форми, в яких всі правильні у даному контексті інтерпретації є одночасно формами і у знахідному і в родовому відмінку, очевидно дасть пустий результат.

(67) [case==acc & case==gen]

Вище наведені запити стосуються інтерпретації сегментів, знайдених в процесі дезамбігуації, тобто в результаті уоднозначення морфо-синтаксичних інтерпретацій. Корпус IPI PAN містить не тільки правильні в даному контексті інтерпретації, а всі інтерпретації даного сегменту, які відомі морфологічному аналізатору.

У деяких випадках може знадобитися доступ до інтерпретацій, відкинутих у процесі уоднозначення, наприклад, тоді, коли треба знайти всі синкретичні форми, що виступили в корпусі, інтерпретаціях родового чи знахідного відмінку, незалежно від того, які з цих інтерпретацій є правильними в даному контексті, або коли треба зігнорувати потенціальні помилкові результати роботи дезамбігуатора. У першому випадку можемо використати наступний запит:

(68) [case~acc & case~gen]

В результаті задання цього запиту будуть знайдені синкретичні форми *go* і *pijanego*, але не форми *jq* і *pijanq*, яким такий синкретизм не притаманний.

Останній вид рівності, доступний в запитах це `~~`. Наведений нижче запит можна використати для знаходження тих форм, про які ще перед уоднозначенням відомо, що вони точно є виключно у знахідному відмінку.

(69) [case~~acc]

Наведена нижче таблиця підсумовує значення окремих операторів рівності.

	у результаті морфологічного аналізу	із врахуванням уоднозначення
щонайменше одна інтерпретація	~	=
кожна інтерпретація	~~	==

Отже, правильними є наступні імплікації:

- [атрибут `~~` значення] → [атрибут `==` значення]
тобто кожна відповідь на запит [атрибут `~~` значення] є також відповіддю на запит [атрибут `==` значення]
- [атрибут `==` значення] → [атрибут `=` значення]
тобто кожна відповідь на запит [атрибут `==` значення] також є відповіддю на запит [атрибут `=` значення]
- [атрибут `=` значення] → [атрибут `~` значення]
тобто кожна відповідь на запит [атрибут `=` значення] також є відповіддю на запит [атрибут `~` значення]

1.6 Обмеження запиту до речення або абзацу

Тексти, які містить Корпус IPI PAN було автоматично поділено на речення і абзаци. Цю інформацію можна використати в запитах, наприклад, обмежити контекст до одного речення.

Щоб обмежити дію запиту, треба дописати до нього ключове слово `within`, а потім `s` або `p`, залежно від того, чи дія має бути обмежена до речення (англ. *sentence*), чи до абзацу (англ. *paragraph*). Це ілюструє наступний приклад запиту про речення, в яких форма *się* виступає після форми лексеми `VAĆ`, але не безпосередньо після неї:

(70) [base=bać] [orth!=się]+[orth=się] within s

1.7 Обмеження запиту з допомогою метаданих

З кожним текстом, що входить до Корпусу IPI PAN, пов'язані так звані метадані, тобто інформація про назву і автора твору, його походження, видавництво і т. п. Частина цієї інформації доступна через програму Poliqar і може бути використана для обмеження дії запиту, наприклад, до текстів конкретного автора або текстів, написаних у певному часовому проміжку.

У даній версії Корпусу IPI PAN є доступні три типи метаданих, яким відповідають п'ять метаатрибутів:

- ім'я і прізвище автора або авторів — атрибут `author`,
- назва твору — атрибут `title`,
- дата написання або публікації твору — атрибути:
 - `created` дата написання твору,
 - `first_published` дата першої публікації,
 - `published` дата публікації тексту, який ввійшов до корпусу.

У більшості випадків лише деякі з цих атрибутів мають значення, наприклад, тоді, коли відома лише дата фактичної публікації, але не відома дата першої публікації і написання, або у випадку коротких нотаток з преси, автор, а іноді і назва яких, не є відомі.

Щоб обмежити дію запиту до окреслених значень метаданих, треба в кінці запиту написати ключове слово `meta`, після якого вписується специфікація значення метаатрибутів. У випадку коли запит обмежений одночасно і до речення чи абзацу, і на підставі метаданих, специфікація обмеження запиту до речення чи абзацу виступає в запиті перед специфікацією метаданих, наприклад:

```
(71) [pos=subst]{6,} within s meta author=Kowalski
```

Значення атрибутів `author` і `title` можна задавати за допомогою регулярних виразів, наприклад, наведений нижче запит можна використати для знаходження форм лексеми WIRUS у творах, назва яких містить одне зі слів *windows* чи *microsoft*.

```
(72) [base=wirus] meta title="windows|microsoft"
```

Специфікація атрибутів `author` і `title` ігнорує величину букв і вони інтерпретуються як частини повних значень, отже, результати наступного запиту будуть походити з творів Pola, Polkowskiego і Rampolskiego (оскільки твори цих авторів є представлені у корпусі):

```
(73) [pos=subst]5 meta author=Pol
```

Прапорці `/X` і `/I` перегукуються із описаними вище прапорцями `/x` і `/i`: прапорець `/X` говорить про те, що дана специфікація атрибуту розуміється як специфікація повної вартості цього атрибуту, а не тільки її частини, а прапорець `/I` говорить про те, що величина букв специфікації зводиться до величини букв значення атрибуту. Це ілюструють наступні приклади:

```
(74) [pos=subst]{5} meta author=Pol/I
пошук обмежено до авторів Pola, Polkowskiego і т. д., але не Rampolskiego
```

```
(75) [pos=subst]{5} meta author="Marek Pol"/X
пошук обмежено до творів Marka Pola
```

```
(76) [pos=subst]{5} meta author=Pol/X
пошук обмежено до творів автора, повна специфікація імені і прізвища якого Pol (або POL, pol і т. д.)
```

- (77) `[pos=subst]{5} meta author=".* Pol"/I`
пошук обмежено до творів автора, прізвище якого *Pol*

По-різному трактуються атрибути `created`, `first_published` і `published`, значеннями яких можуть бути виключно дати. У випадку атрибутів цього типу неможливим є вживання регулярних виразів, але можна вжити знаки `< i >` з метою окреслення часового проміжку, з якого має походити даний твір, наприклад:

- (78) `[pos=subst]{5} meta created>1950`
пошук обмежено до творів написаних після 1950 року

Умови, що стосуються метаданих можна об'єднувати за допомогою операторів `&`, `|` і `!`:

- (79) `[pos=subst]5 meta created>=1951 & created<=1960`
пошук обмежено до творів, написаних в 1951–1960 роках
- (80) `[pos=subst]5 meta published>1900 & author!=Sienkiewicz`
пошук обмежено до творів, виданих після 1900 року, за винятком текстів Sienkiewicza
- (81) `[pos=subst]5 meta (author=sienkiewicz & title=potop)`
`| (author=zeromski & title=przedwiośnie)`
пошук обмежено до Potopu Sienkiewicza і Przedwiośnia Żeromskiego

У даній версії корпусу багато творів не містять повних метаданих. Як наслідок, в результаті наведених вище запитів, у яких подані обмеження за метаатрибути, будуть знайдені лише ті твори, для яких для відповідних метаатрибутів, задані відповідні значення. Отже, результатом задання першого з наведених нижче запитів буде знаходження значно меншої кількості результатів, ніж у випадку задання другого запиту.

- (82) `[pos=subst] meta published>1900 | published<=1900`
- (83) `[pos=subst]`

У випадку, коли даному метаатрибуту, не є приписане відповідне значення, приймається, що його значенням є пуста послідовність, тобто `""`. Отже, наступний запит є рівнозначний до другого з попередніх запитів.

- (84) `[pos=subst] meta published>1900 | published<=1900`
`| published=""`

1.8 Вирівнювання результатів

У частині запиту, що стосується форм, тобто перед кваліфікаторами `within` і `meta`, можна використовувати знак вирівнювання `^`, наприклад:

- (85) `[pos=adj & case=nom]^ [pos=subst & case=nom]^`

У результаті задання вищенаведеного запиту буде висвітлено чотири стовпці, що містять відповідно: лівий контекст, тобто фрагмент тексту, що йде безпосередньо перед знайденим фрагментом, лівий заданий фрагмент, тобто фрагмент тексту, що відповідає частині запиту перед знаком `^` (тут: послідовність прикметників у називному відмінку), правий контекст (тут: послідовність іменників у називному відмінку), як це показано на Рис.1.

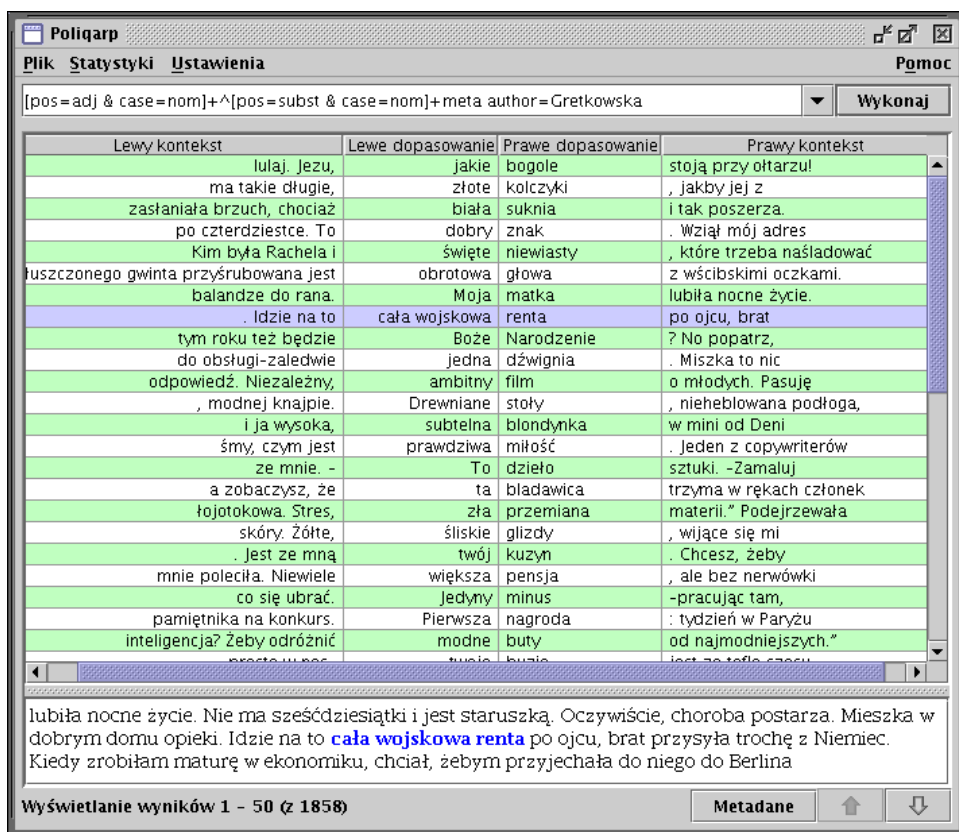


Рис. 1: Вирівнювання результатів в графічній версії

2 Poliqarp

Цей пункт містить опис трьох версій інтерфейсу програми Poliqarp, починаючи від найпростішої мережевої версії (п. 2.1), графічної версії (п. 2.2), призначеної для більшості користувачів, і закінчуючи текстовою версією, яка працює під операційною системою LINUX (п. 2.3) та має найбільші можливості і швидкодію, але найменш привабливий інтерфейс.

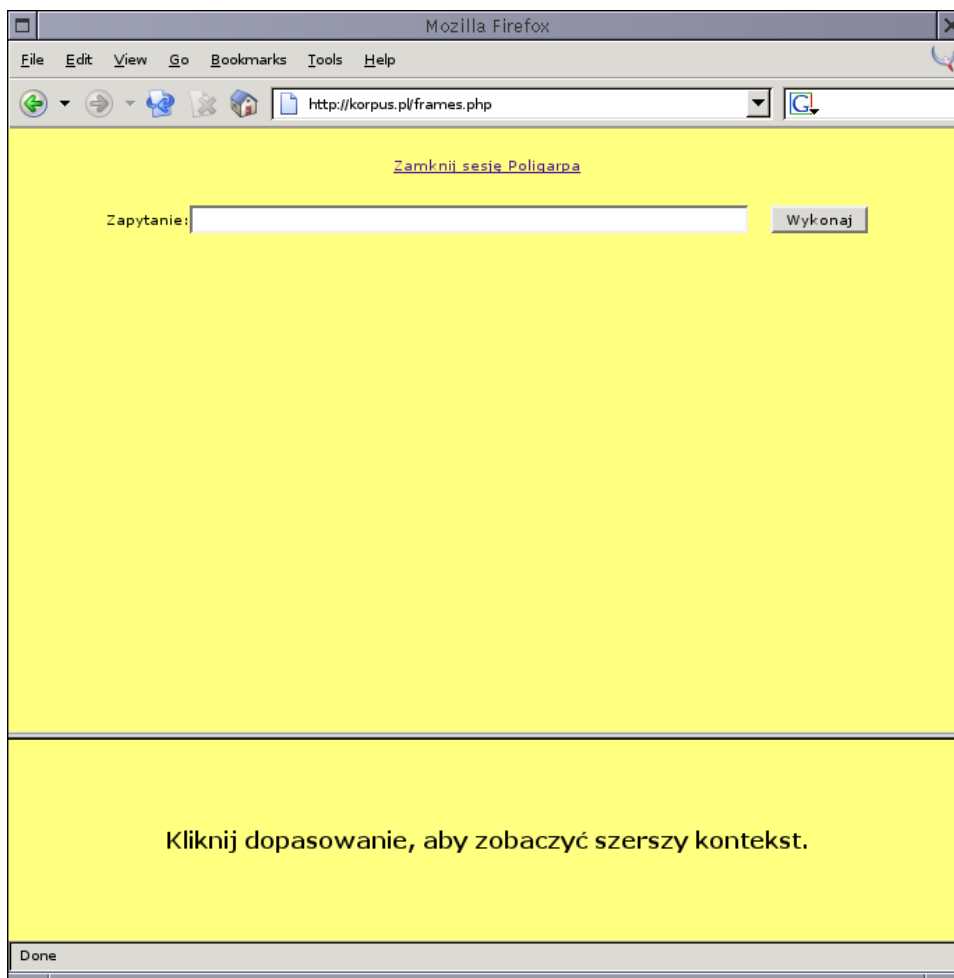


Рис. 2: Інтернет-версія безпосередньо після вибору корпусу

2.1 Інтернет-версія

У мережевій версії Корпус ІРІ РАН доступний за адресою `korpus.pl`. На даний момент корпус є доступний без обмежень, хоча, можливо, буде впроваджено обмеження, якщо популярність корпусу буде більшою, ніж це дозволяє потужність сервера, що обслуговує корпус.

Хоч функціональність актуальної Інтернет-версії є досить обмежена порівняно із графічною і текстовою версіями, у ній немає обмежень на систему запитів, описану в п. 1. Безпосередньо після входження на сторінку `korpus.pl` і після обрання корпусу висвітлюється рядок запитів, як це показано на Рис. 2.

У цьому рядку, згідно з описом в п. 1, можна вписати будь-які запити, а потім натиснути **Enter** або натиснути на кнопку **Wykonaj**. Результати будуть висвітлені в трьох або чотирьох стовпцях, залежно від того, чи присутній у запиті знак вирівнювання, див. Рис. 3. Натискання на одному з результатів, а точніше на центральній частині результату, що відповідає контекстові запиту, приведе до висвітлення ширшого контексту в нижній частині сторінки, як показано на Рис. 4.

Результати пошуку показуються групами по 25 результатів. Щоб перейти до наступних 25 результатів, треба клікнути на напис **Наступні 25**, а щоб повернутись до попередніх результатів — напис **Попередні 25**.

Єдиною додатковою функцією, доступною в Інтернет-версії, є можливість сортування результатів за будь-якими із стовпців, за алфавітом *a fronte* (від початку значення) або *a tergo* (від кінця значення), в порядку зростання або

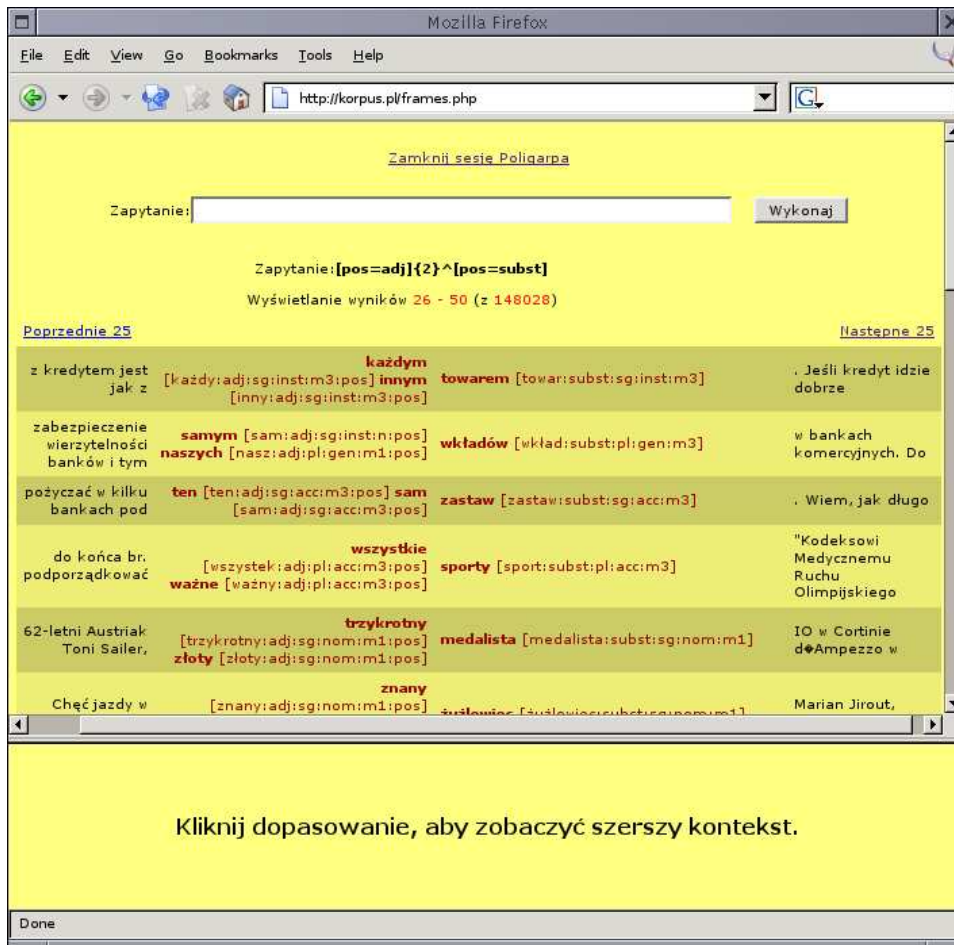


Рис. 3: Інтернет-версія: результати пошуку

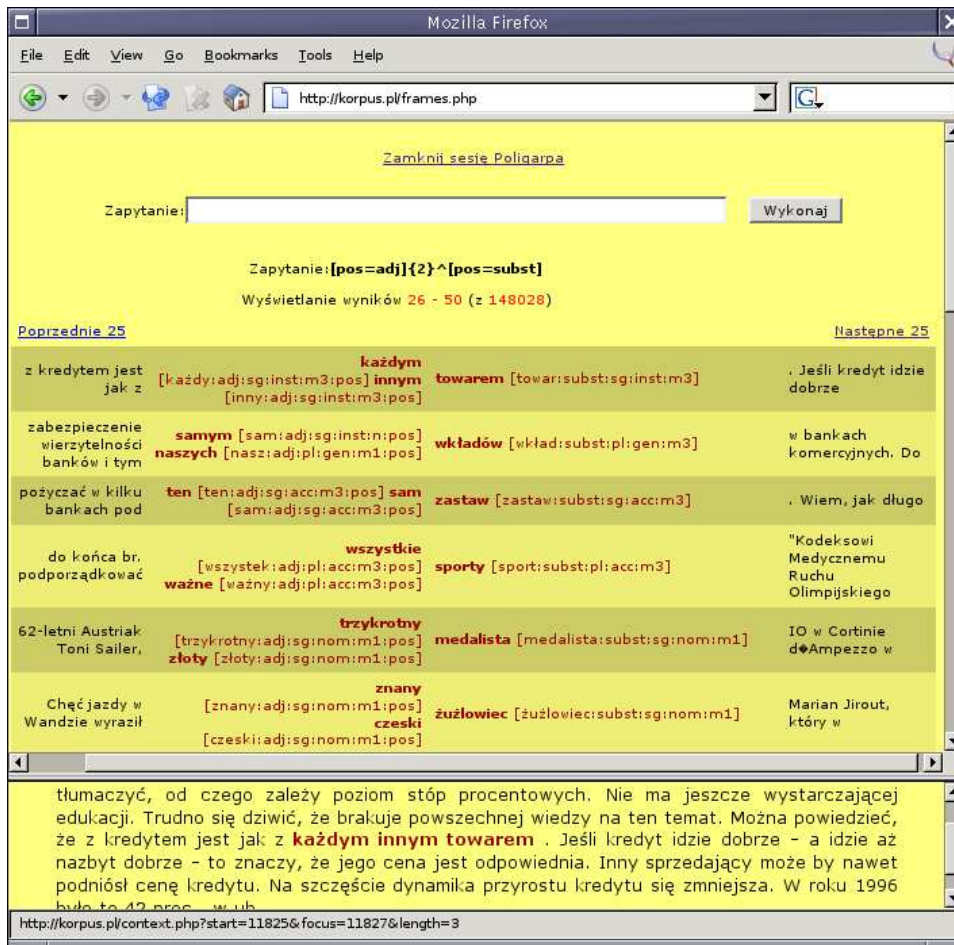


Рис. 4: Інтернет-версія: ширший контекст

спадання.

Інтернет-версія програми Poliqar є найчастіше змінюваною версією, тому можливо, що в момент, коли Ви читаєте цю публікацію, її робота буде істотно відрізнитися від поданого тут опису.

2.2 Графічна версія

Основною версією програми Poliqar є графічна. Крім меню, вікно графічної версії містить рядок запитів, верхня частина вікна служить для відображення результатів, а нижня містить детальнішу інформацію про знайдений результат: ширший контекст, як на Рис. 1 на стор. 15, або метадані.

Безпосередньо після запуску програми, рядок запитів і обидві частини є пустими, як на Рис. 5.



Рис. 5: Графічна версія безпосередньо після запуску

Щоб зробити можливим пошук інформації в корпусі, треба спочатку завантажити файл з цим корпусом, вибравши в меню **Файл** команду **Відкрити корпус...**, а потім вказати де знаходиться каталог з Корпусом IPI PAN (на CD-ROM або на жорсткому диску, залежно від того, як був заінстальований корпус). Подібним чином можна змінити завантажений корпус, не виходячи з програми. При читанні корпусу запам'ятовується інформація про його місцезнаходження — при нових запусках програми Poliqar досить вибрати пункт **Останній корпус** в меню **Файл**, або натиснути комбінацію клавіш **Ctrl-Alt-1** для зчитування першого з запам'ятованих корпусів, натомість, щоб прочитати другий запам'ятований корпус треба натиснути **Ctrl-Alt-2** і т. д.

Після зчитування корпусу можна вписати запит у рядку запитів і запустити пошук натисканням клавіші **Enter** або кнопки **Виконати**. Якщо запит не містив знаку вирівнювання результатів, то в верхній частині вікна будуть висвітлені результати в трьох стовпчиках, що містять відповідно лівий контекст, заданий фрагмент і правий контекст, як показано на Рис. 6, причому ширину стовпців можна змінювати, перетягуючи мишею границі заголовків стовпців. У випадку задання запиту зі знаком вирівнювання **^** буде висвітлено чотири стовпці, як описано у п. 1.8.

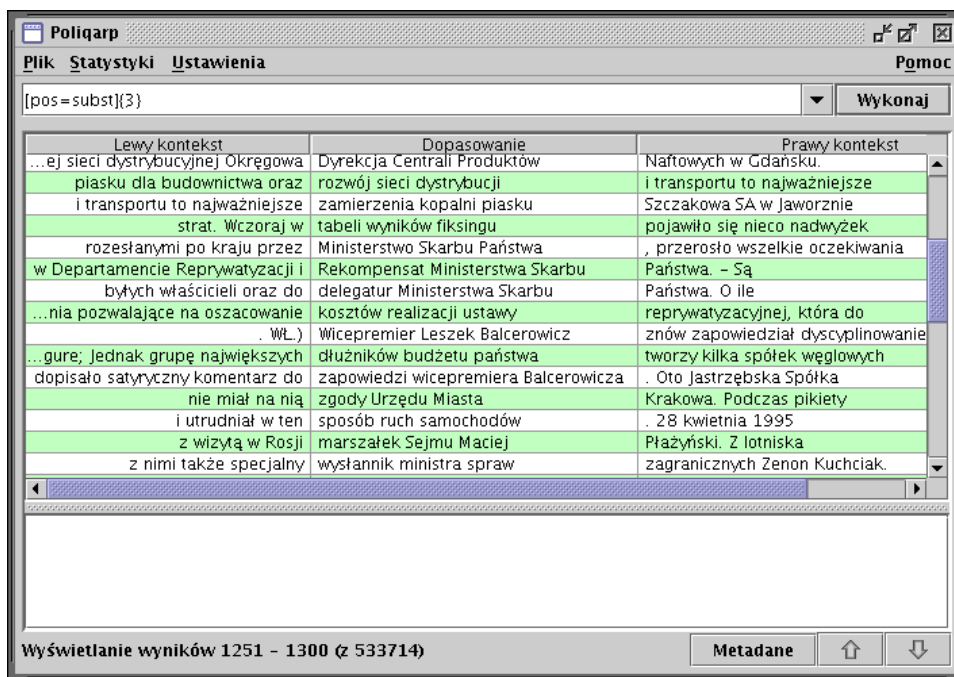


Рис. 6: Результати запиту без вирівнювання

Програма Poliqarp зберігає історію заданих у даній та попередніх сесіях запитів. На історію запитів можна вийти через малу кнопку, що знаходиться між рядком запитів і кнопкою Виконати. У результаті запиту у верхній частині вікна Poliqarp буде показано перші 50 результатів. Щоб подивитися наступні результати, треба натиснути на стрілку, що знаходиться у правому нижньому куті вікна. Повернутися до попередніх результатів можна, натиснувши на стрілку вгору, що знаходиться зліва.

Щоб побачити ширший контекст вибраного результату, треба натиснути на відповідний рядок у верхній частині вікна. У результаті рядок буде виділено іншим кольором, а ширший контекст буде показано у нижній частині вікна. В нижній частині представлено також інформацію про текст, з якого взято даний фрагмент, чи так звані метадані, а для переходу між режимом представлення контексту і режимом представлення метаданих служить кнопка **Метадані/Контекст** внизу вікна.

Результати пошуку можна сортувати по даних кожного стовпчика. У найпростішому випадку, щоб відсортувати результати в порядку зростання, треба клікнути лівою клавішею миші на заголовку відповідного стовпчика, наприклад, стовпчика **Заданий фрагмент** (у випадку запиту без знаку вирівнювання). Повторне клікання приведе до сортування результатів по даних того ж стовпчика, але в порядку спадання, а наступне — знову сортування в порядку зростання.

Також є можливість сортування *a tergo*, тобто від кінця значення, і в зростаючому і в спадаючому порядках. Ця функція доступна в меню, яке появляється після клікання правою клавішею миші на заголовку відповідного стовпчика, як це показано на Рис. 8, який ілюструє сортування *a tergo* в порядку зростання по значеннях стовпчика **Лівий заданий фрагмент**.

Спосіб сортування також можна встановити в меню **Налаштування**, вибравши **Опції...**, а потім закладку **Сортування**. Вибрані таким чином значення будуть сортуватися при представленні результатів чергових запитів і будуть збережені для використання у майбутніх сесіях.

Серед інших опцій, які можна використовувати в меню **Налаштування** → **Опції...** є ширина контексту, що рахується у сегментах, який буде висвітлено

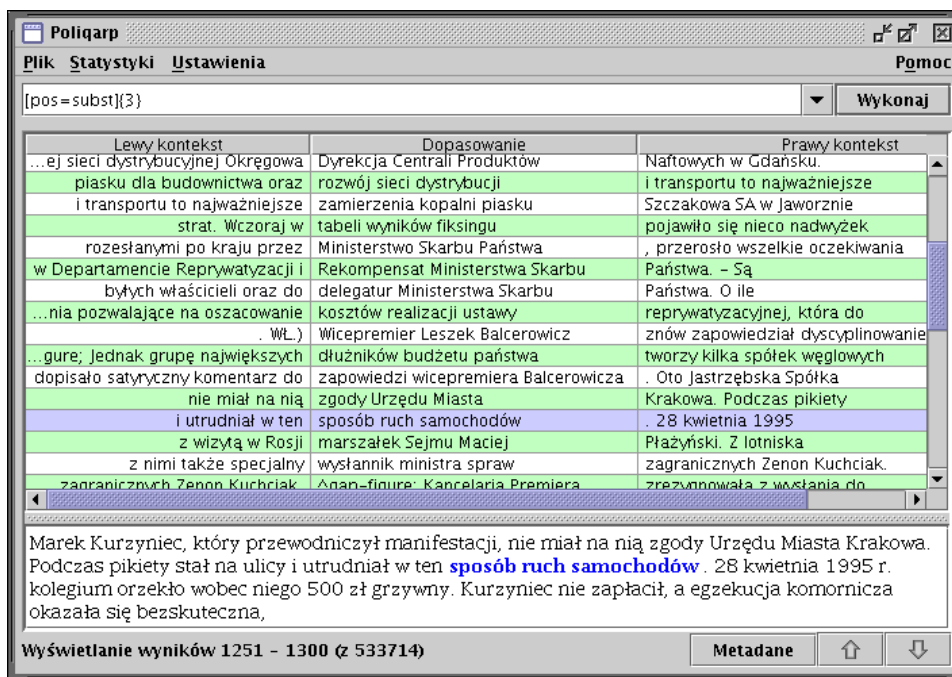


Рис. 7: Ширший контекст результату пошуку

для кожного знайденого результату запиту: як ширина лівого і правого контекстів у верхній частині вікна, так і більшого контексту в нижній частині вікна (закладка Контекст).

Важливою рисою програми Poliqarp є можливість висвітлення не лише словоформ, що виступають у корпусі, а також їх початкових форм (лем) і морфосинтаксичних тегів. У сучасній версії програми опцію висвітлення словоформ, лем і тегів можна задати окремо для стовпців з заданим фрагментом і окремо для стовпців з контекстом. Наприклад, ефектом налаштувань з Рис. 9 буде форматування результатів показане на Рис. 10.

Як було сказано вище (стор. 9), графічна версія програми Poliqarp дає можливість задання так званих аліасів, чи скорочень для альтернатив значень даного атрибуту. Для редагування аліасів, а також для їх додавання, служить закладка Аліаси в Налаштування → Опції. У даній версії програми аліаси, що були додані під час однієї сесії, не зберігаються для використання під час наступних сесій.

Наступна закладка Прапорці дає можливість окреслити спосіб інтерпретації поданих користувачем значень атрибутів у частині запиту, що стосується шуканих форм і в частині запиту, що стосується метаданих. За замовчуванням запит типу [pos=ppro] знайде ті сегменти, граматичний клас яких є ppro, отже результат такого запиту буде пустим. Щоб знайти різні займенникові форми, тобто такі, граматичний клас яких rpron12 або rpron3, можна використати прапорець /x. Щоб змінити такий спосіб поведінки програми за замовчуванням, щоб вона інтерпретувала всі запити про форми так, як при увімкненому прапорці /x, треба дезактивувати значення опції Тільки цілі слова у стовпці Запит в закладці Прапорці.

Аналогічно можна змінити інтерпретацію специфікації метаатрибутів. За замовчуванням специфікація метаатрибуту має відповідати принаймні частині значення метаатрибуту, але це можна змінити для потреб конкретного запиту з допомогою прапорця. Щоб змінити це значення на постійне, треба активувати опцію Тільки цілі слова у стовпці Метадані на закладці Прапорці.

Подібно, за допомогою встановлення опції Ігнорувати величину букв, можуть бути змінені стандартні інтерпретації чутливості до величини букв в обидвох

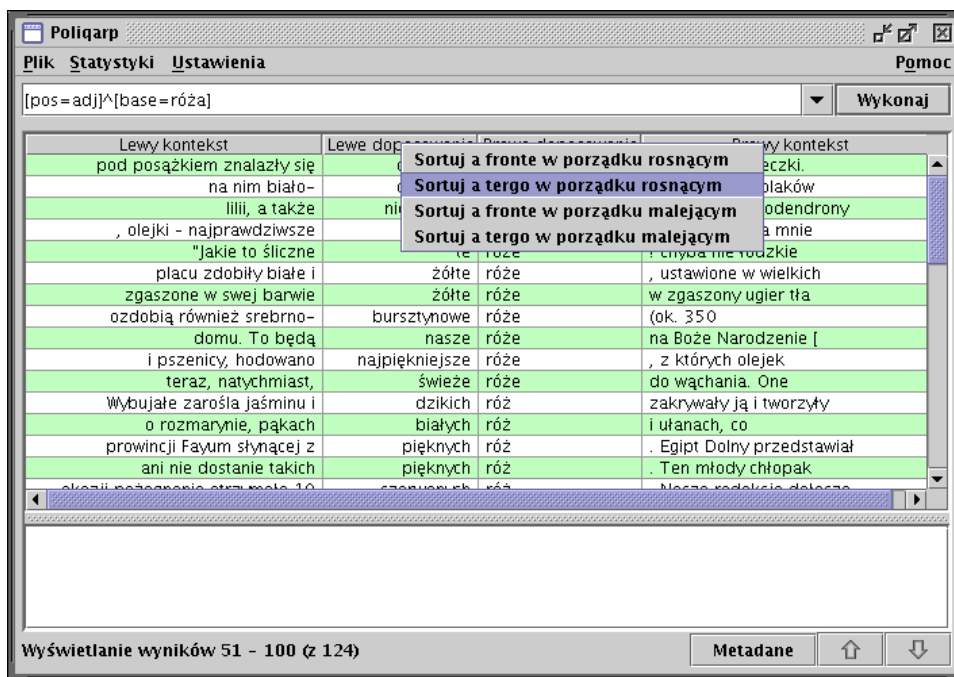


Рис. 8: Стортування *a tergo* в порядку зростання

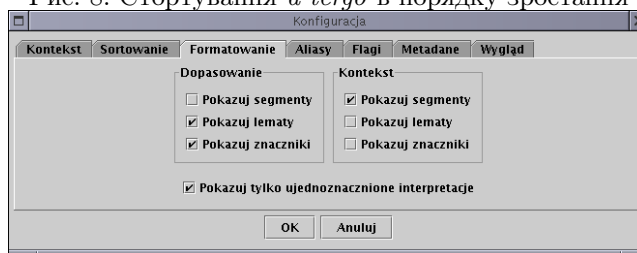


Рис. 9: Опції форматувannya результатів

частинах запитів.

Наступна закладка в меню Налаштування → Опції... це закладка Метадані. З її допомогою можна обмежити майбутні запити до наперед заданих значень мегаатрибутів, наприклад, до творів конкретного автора або творів, опублікованих у заданому часовому проміжку. Остання закладка, Вигляд, дозволяє встановити розмір шрифту.

Крім підменю Файл і підменю Налаштування, в головному меню програми Poliqarp доступне також підменю Статистика. У даній версії програми вибір цього підменю висвітлює основну чисельну інформацію про завантажений корпус. Наприклад, підкорпус Корпусу IPI PAN, до якого належить інформація, показана на Рис. 11, містить понад 56 мільйонів позицій, що відповідають 729 тисячам різних сегментів, що є формами понад 355 тисяч різних початкових форм. Кількість різних морфо-синтаксичних тегів, під якими розуміємо $k1:k2:k3:\dots:k_n$ (див. стор. 10) в цій версії корпусу 1150.

Для запису результатів використовується позиція Записати результати... в меню Файл, а для завершення роботи з програмою — позиція Вихід у тому ж меню.

2.3 Текстова версія

Текстова версія програми Poliqarp призначена для операційної системи GNU/Linux, що працює на комп'ютерах класу PC.

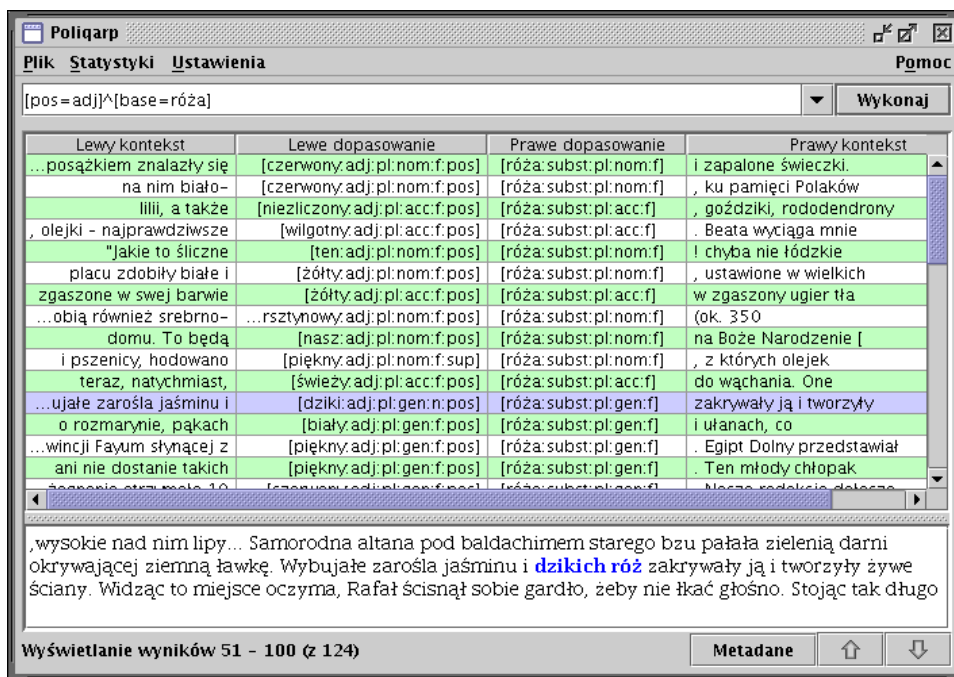


Рис. 10: Результати форматування при опціях з Рис. 9

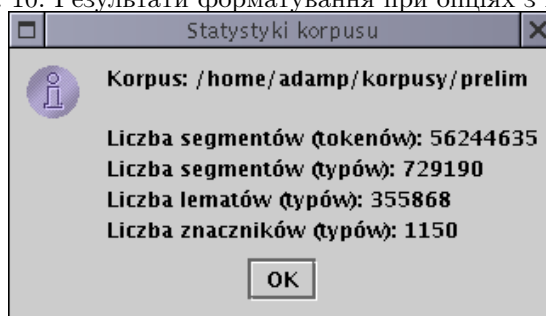


Рис. 11: Кількісна інформація про відкритий корпус

Програму можна запустити за допомогою команди `poliqarp korpus`, де `korpus` це назва корпусу (перша частина назв файлів `wstepny.cfg` і `wstepny.poliqarp.corpus.image`), включно зі шляхом до цього корпусу. Наприклад, якщо корпус знаходиться в каталозі `./korpus/` і складається з файлів типу `wstepny.cfg`, `wstepny.poliqarp.chunk.image` і т.п., програму можна запустити наступним чином:

(86) `$ poliqarp korpus/wstepny`

В результаті виконання попередньої команди, командний рядок оболонки буде змінено на командний рядок програми Poliqarp, наприклад:

(87) `KORPUS/WSTEPNY>`

Командний рядок програми Poliqarp служить для вводу запитів та інших команд. Щоб задати запит, треба просто вписати його у командний рядок і натиснути клавішу `Enter`. Наприклад, в результаті запиту показаного в (88) буде виведено результати як на Рис. 12.

(88) `KORPUS/WSTEPNY> [pos=adj]^[base="róża|stokrotka"]`

Якщо користувач не задав інших опцій — результати запиту генеруються у форматі HTML, а для їх відображення використовується програма Links.

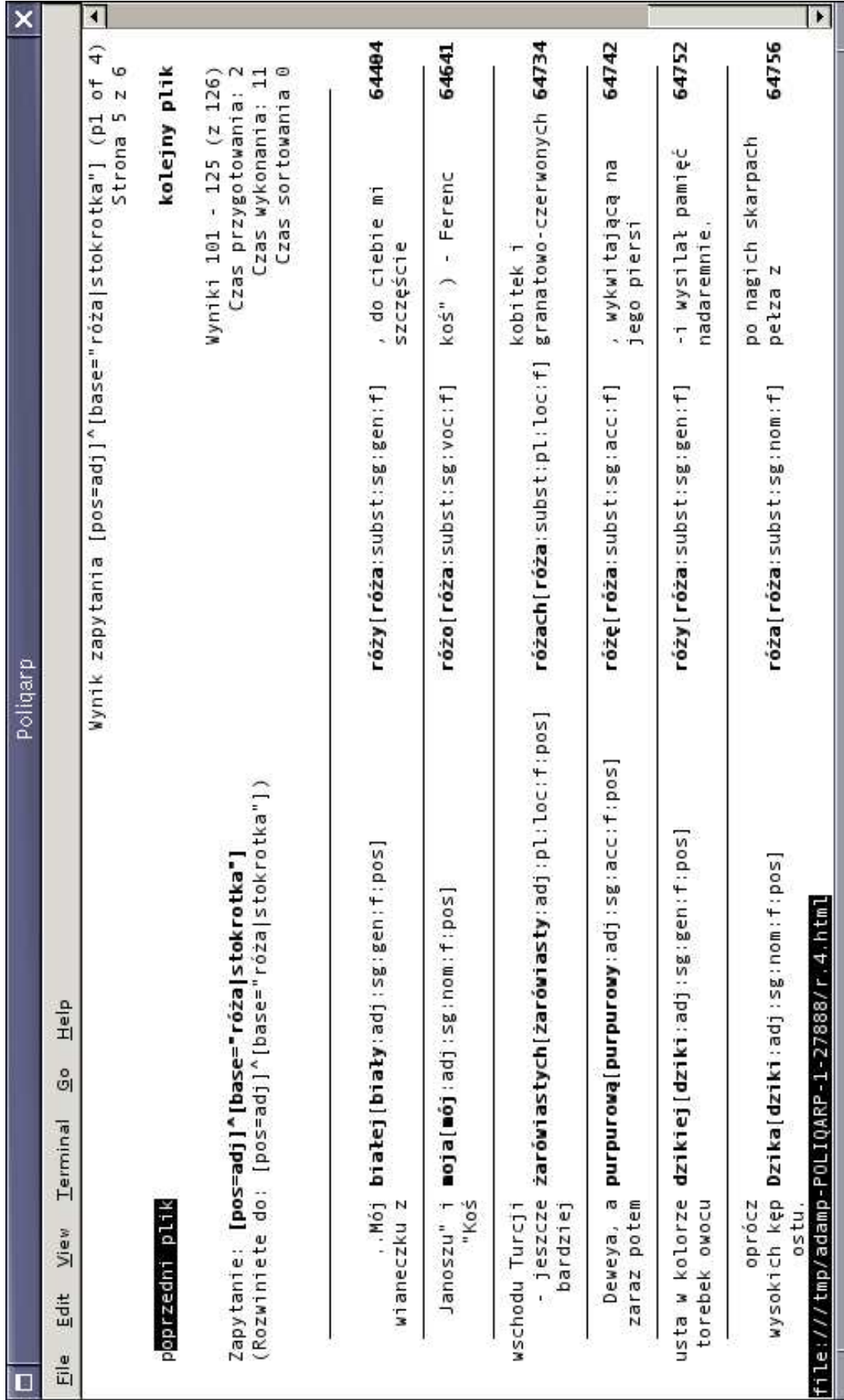


Рис. 12: Результат запиту в текстовій версії (конфігурація за замовчуванням)

2.3.1 Редагування запитів у командному рядку

Подібно як і в випадку графічної версії програми, текстова версія також зберігає певну кількість попередніх запитів (а також команд, викликаних з командного рядка). Доступ до цих команд і запитів можна отримати за допомогою клавіші зі стрілкою вгору і зі стрілкою вліво.

В обмеженому обсязі працює також пошук в історії запитів: у результаті натиснення комбінації клавіш `Ctrl-r` і вводу послідовності знаків буде знайдено (в історії запитів, а не в корпусі!) останній запит, що містить цю послідовність знаків.

Запит можна редагувати за допомогою клавіш зі стрілками вліво і вправо, клавіші `Backspace` а також наступних комбінацій клавіш, відомих з програми `Emacs`:

клавіша	ефект
<code>Ctrl-b</code>	перехід курсора на один знак вліво
<code>Ctrl-f</code>	перехід курсора на один знак вправо
<code>Ctrl-a</code>	перехід курсора на початок рядка
<code>Ctrl-e</code>	перехід курсора на кінець рядка
<code>Esc-b</code>	перехід курсора на одне слово вліво
<code>Esc-f</code>	перехід курсора на одне слово вправо
<code>Ctrl-d</code>	видалення знаку в позиції курсора
<code>Esc-d</code>	видалення знаків від позиції курсора до кінця слова
<code>Esc-Backspace</code>	видалення знаків від початку слова до знаку перед курсором
<code>Ctrl-k</code>	видалення знаків від позиції курсора до кінця рядка
<code>Esc-4 Ctrl-f</code>	як чотирикратне натискання <code>Ctrl-f</code> , тобто перехід курсора на чотири знаки вправо
<code>Esc-2 Esc-d</code>	як подвійне натискання <code>Esc-d</code>
	і т.д.

Натомість клавіші `Del`, `Home` і `End` не працюють в очікуваний спосіб.

В момент задання запиту (через натиснення клавіші `Enter`) курсор може знаходитись у довільному місці командного рядка.

2.3.2 Конфігурація програми

Роботу текстової версії програми `Poliqar` можна істотно модифікувати, редагуючи конфігураційний файл `.poliqarp_config`, що знаходиться у домашньому каталозі користувача. Найпростіший спосіб створення такого конфігураційного файлу — виклик з командного рядка команди `/dump-config`. Наслідком цього буде запис в домашньому каталозі користувача файлу

```
.poliqarp_config.new,
```

що містить опції текстової версії програми `Poliqar` за замовчуванням. Щоб цей файл став обов'язковим конфігураційним файлом, треба змінити його назву на `.poliqarp_config`.

Конфігураційний файл *de facto* є послідовністю команд для програми `Poliqar`, які виконуються при старті програми. Кожну з цих команд, можна також задати в будь-який момент з командного рядка програми `Poliqar`, ввівши перед нею `/`.

Наприклад, щоб змінити довжину (в сегментах) правого контексту для даної сесії, треба виконати наступну команду:

```
(89) KORPUS/WSTEPNY> /set right-context 20
```

Цій команді відповідає наступний запис у конфігураційному файлі:

```
(90) set right-context 20
```

Аналогічно можна змінити довжину лівого контексту, або виконавши команду в (91) (зміна буде стосуватися даної сесії), або додавши до конфігураційного файлу запис (92) (зміна буде стосуватися усіх сесій).

```
(91) KORPUS/WSTEPNY> /set left-context 20
```

```
(92) set left-context 20
```

Змінні `right-context` і `left-context` це лише дві з багатьох змінних текстової версії програми `Poliqar`, які можна модифікувати за допомогою команди `set`. Короткий опис всіх змінних можна отримати за допомогою виконання з командного рядка команди `/desc`⁶, а актуальні значення всіх змінних — `/show`. Якщо ці команди будуть викликані з аргументами, що є назвами змінних, то буде виведена відповідна інформація лише для цих змінних. Наприклад, щоб отримати інформацію про актуальні значення змінних `right-context` і `left-context`, треба виконати команду:

```
(93) KORPUS/WSTEPNY> /show right-context left-context
```

За замовчуванням, як показано на Рис. 12, у стовпцях з лівим і правим контекстами виводяться виключно сегменти, тобто форми, що виступили в тексті, а в стовпці (або стовпцях) з заданим фрагментом — сегменти, початкові форми і морфо-синтаксичні теги. Тип інформації, що виводиться, можна змінювати, надаючи відповідні значення змінним `cl-x` (щоб змінити лівий фрагмент), `ml-x` (лівий заданий контекст), `mr-x` (правий заданий фрагмент або, якщо у запиті не було знаку вирівнювання, цілий заданий фрагмент) і `cr-x` (правий контекст). Значенням кожної з цих змінних має бути послідовність букв. Якщо серед цих букв є буква `o` (як `orth`), у даному стовпці буде виведена словоформа сегменту. Якщо серед них є буква `b` (як `base`), буде виведена початкова форма. А якщо серед них є буква `t` (як `tag`), буде виведено морфо-синтаксичні теги, що відповідають даному контексту (тобто ті, що були приписані слову після дезамбігуації), а у випадку букви `T` — всі теги приписані даному сегменту морфологічним аналізатором. Наприклад, якщо у конфігураційному файлі є рядки показані в (94), результат запити буде виглядати як на Рис. 13.

```
(94) set cl-x bo
      set ml-x bt
      set mr-x bt
      set cr-x o
```

У даній текстовій версії результати пошуку в корпусі генеруються в форматі HTML-файлів, кожен з яких, можливо, окрім останнього, містить максимальну кількість результатів, що є значенням змінної `hits-per-page`. Ці результати виводяться за допомогою програми, назва якої є значенням змінної `pager`. Наприклад, виконання наведених нижче команд, з командного рядка зумовить виведення результатів наступних запитів за допомогою програми `Mozilla`, по 50 результатів на сторінці.

```
(95) KORPUS/WSTEPNY> /set hits-per-page 50
```

```
(96) KORPUS/WSTEPNY> /set pager mozilla
```

Формат виводу результатів можна досить детально модифікувати за допомогою змінних `result-format` і `match-format`: перша з них описує формат HTML цілої сторінки з результатами, а друга — формат окремого рядка в таблиці з результатами. Значення цих змінних не повинні модифікуватися користувачами, які не знають досконало формату HTML. Детальний опис цих змінних виходить за рамки цієї публікації.

⁶Як і інші команди текстової версії, команду `desc` також можна записати в конфігураційному файлі, що виводитиме короткий опис змінних при кожному запуску програми `Poliqar`.

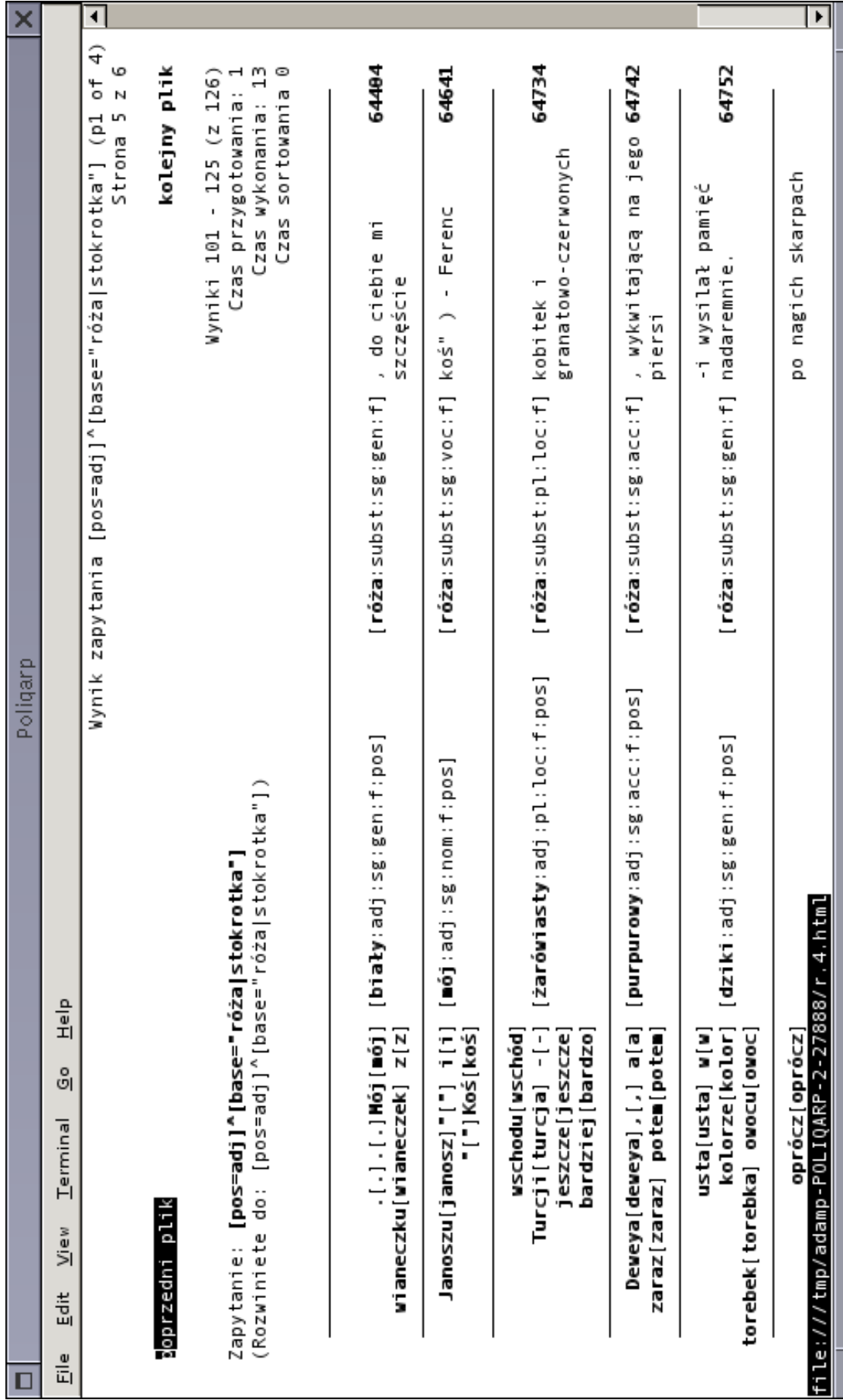


Рис. 13: Форматування результатів при опціях в (94)

Варто згадати про такі змінні, як `input-encoding` і `output-encoding`. Їх значення описують відповідно кодування входу і виходу терміналу, в якому запускається текстова версія. Як правило не варто змінювати значення цих змінних, які є встановлені за замовчуванням (`latin2`).

Як і у випадку графічної версії, в текстовій версії також можливим є сортування результатів по кожному із стовпчиків таблиці з результатами, в порядку зростання або спадання, *a fronte* (за алфавітом від початку) або *a tergo* (за алфавітом з кінця). Порядок сортування описується за допомогою змінної `sort-by`, значенням якої є послідовність знаків, впорядкована згідно із конфігурацією (налаштуваннями сортування). Наприклад, команда в (97) зумовить, сортування результатів наступних запитів (*a fronte*, у зростаючому порядку) по правому контексту.

```
(97) KORPUS/WSTEPNY> /set sort-by r
```

Кожному стовпцю відповідає інша буква:

буква	стовпець сортування
<code>l</code>	лівий контекст
<code>r</code>	правий контекст
<code>n</code>	ліва частина заданого фрагменту
<code>m</code>	права частина заданого фрагменту

Запис в значеннях змінної `sort-by` малих букв зумовить сортування *a fronte*, а запис відповідних великих букв зумовить сортування *a tergo*. Додатково перед кожною буквою може стояти знак `+` для сортування в порядку зростання або `-` для сортування в порядку спадання. Відсутність знаку перед буквою інтерпретується як знак `+`. Наприклад, команда (98) зумовить сортування результатів у порядку спадання по лівому контексту, *a tergo*.

```
(98) KORPUS/WSTEPNY> /set sort-by -L
```

Якщо в значеннях змінної `sort-by` задано декілька стовпчиків, то сортування відбувається в першу чергу за значеннями першого стовпчика, потім для тих самих значень у першому стовпчику — за значеннями другого поданого стовпчика і т.д. Наприклад, наведена нижче команда зумовить сортування результатів запитів в порядку зростання за стовпчиком з правою частиною заданого фрагменту, а потім, враховуючи результати цього сортування, в порядку спадання по значеннях стовпчика з лівою частиною заданого фрагменту, як це показано на Рис. 14.

```
(99) KORPUS/WSTEPNY> /set sort-by m-n
```

Результати останнього запиту запам'ятовуються аж до виконання наступного запиту і можуть бути ще раз посортовані і виведені в новому порядку. Для цього служать команди `sort` і `view`. Наприклад, щоб переглянути результати останнього запиту, але посортовані по лівому контексту, треба виконати наступну команду:

```
(100) KORPUS/WSTEPNY> /set sort-by l
```

```
(101) KORPUS/WSTEPNY> /sort
```

```
(102) KORPUS/WSTEPNY> /view
```

У випадку всіх описаних вище змінних замість їх повних назв можна вживати скорочення, причому скороченням може бути початок назви довільної довжини, наскільки лише він дозволяє однозначно ідентифікувати дані змінні. Наприклад, замість команди `set c1-x bo` в (94) можна подати команду `set c1 bo`, в якій назва змінної `c1-x` є однозначно скорочена до `c1`, але не можна

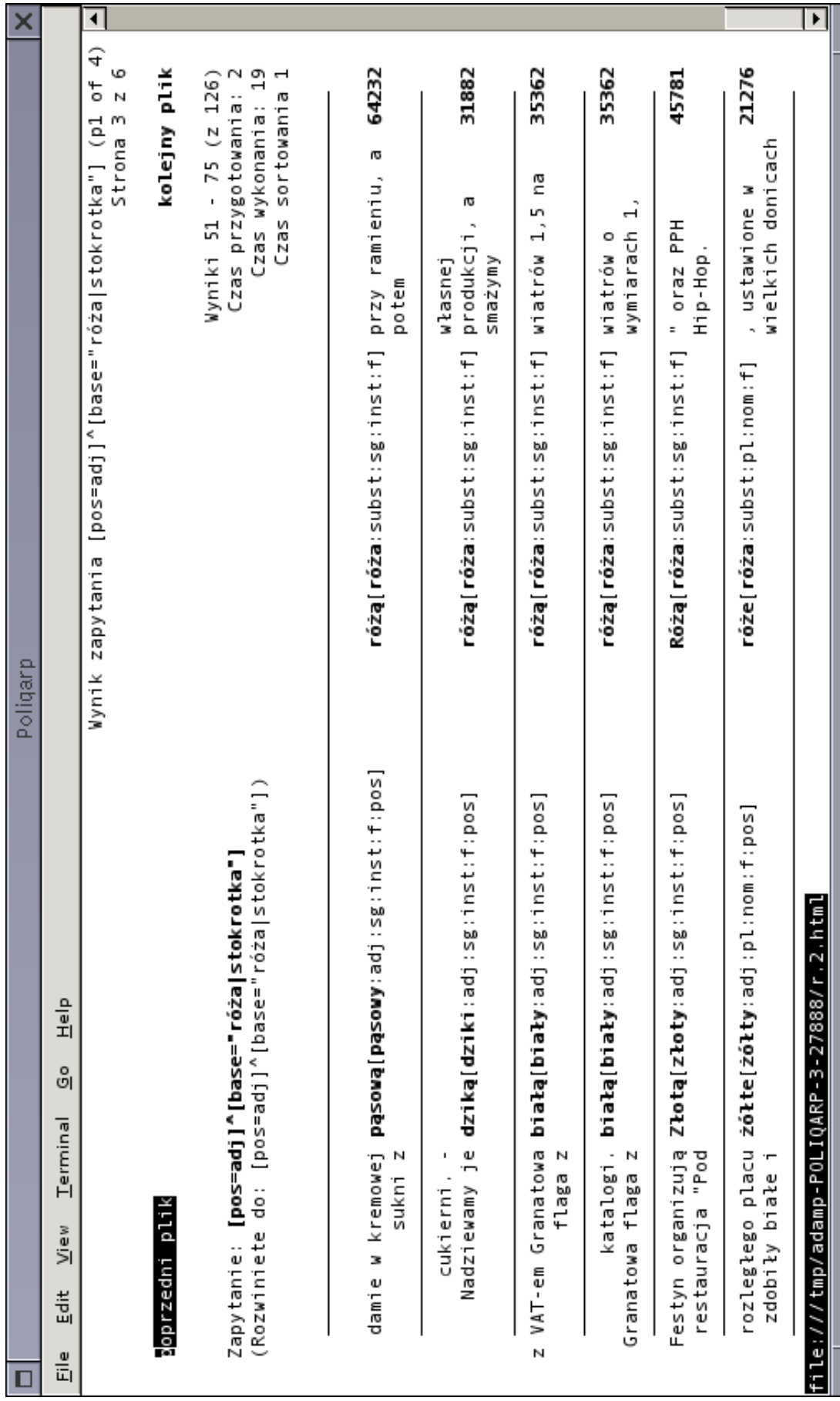


Рис. 14: Сортування результатів за специфікацією в (99)

задати команду `set c bo`, оскільки скорочення `c` не є однозначне: невідомо, чи воно відноситься до `cl-x` чи до `cr-x`.

Як і у випадку графічної версії, можна переглядати дані про текст, з якого взято даний результат пошуку. Як видно з Рис. 12–14, в останньому стовпці таблиці з результатами знаходиться певне число — воно є номером тексту, до якого належить дана цитата. Щоб отримати інформацію про цей текст, треба в командному рядку ввести `/meta <number>`, наприклад, `/meta 64232`.

У текстовій версії можна також оголошувати додаткові аліаси. Для цього служить команда `alias`. Наприклад, щоб можна було вживати напис `ppron` як скорочення для `ppron12` і `ppron3`, треба у конфігураційний файл вписати наступний рядок:

```
(103) alias prpron = prpron12 prpron3
```

Звичайно, як і у випадку інших команд, її можна виконати і з командного рядка, поставивши знак `/` перед нею. Команда `alias` без аргументів зумовить вивід оголошених аліасів, а команда `unalias` з одним або більшою кількістю аргументів, що є назвами аліасів, зумовить знищення цих аліасів.

Щоб закінчити роботу з програмою `Poliqar`, треба виконати команду `exit`. Без виходу з програми можна з допомогою команди `open`, змінити корпус. Наприклад, щоб завантажити корпус `frek`, що знаходиться в каталозі `/home/adamp/korpus/`, треба виконати команду

```
(104) KORPUS/WSTEPNY> /open "/home/adamp/korpus/frek"
```

Стислу інформацію про доступні команди текстової версії програми `Poliqar` можна отримати, виконавши команду `help`, а команда `stat` зумовить виведення основної інформації про відкритий корпус.

2.3.3 Системні змінні

Текстова версія має дві мовні версії: польську і англійську. Мова вибирається автоматично на підставі системної змінної `LANG`. Якщо значенням цієї змінної є `pl_PL`, то це польська версія; в протилежному випадку — англійська версія. Інші стандартні системні змінні, значення яких можуть вплинути на роботу програми, це `HOME`, `TMP`, `TEMP` і `USER`. Їх роль детально описана у файлі `README.pl.txt` в каталозі `linux` на CD-ROM.

A GNU Free Documentation License

Версія 1.2

Листопад 2002

Copyright (C) 2000,2001,2002 Free Software Foundation, Inc.
59 Temple Place, Suite 330, Boston, MA 02111-1307 USA
Кожен має право копіювати і поширювати копії даної Ліцензії
за умови дослівного збереження її тексту.

A.1 ПРЕАМБУЛА (PREAMBLE)

Мета даної Ліцензії — зробити посібник, збірник текстів чи інший функціональний і корисний документ "відкритим": тобто гарантувати кожному ефективно право копіювати і розповсюджувати його, зі змінами чи без, на комерційній чи некомерційній основі. Дана Ліцензія також зберігає за автором чи видавцем право отримати визнання за свою роботу, не несучи відповідальності за зміни, зроблені іншими.

Дана Ліцензія відноситься до категорії "copyleft", яка означає, що роботи, що є що похідними з документу, повинні бути відкритими в такому ж розумінні. Дана Ліцензія доповнює GNU General Public License, яка також є ліцензією "copyleft", розробленою для відкритого програмного забезпечення.

Дана Ліцензія розроблена для посібників ("керівництв") до відкритого програмного забезпечення, оскільки відкрите програмне забезпечення вимагає відкритої документації: відкрита програма повинна супроводжуватися посібником, щодо якого діє така ж відкритість як і до самої програми. Але дія даної Ліцензії не обмежується тільки посібниками до програмного забезпечення і може бути використана для будь-якої текстової роботи, незалежно від теми чи факту публікації даної праці в вигляді книги. Дану Ліцензію рекомендується використовувати, головним чином, для робіт навчального та довідкового характеру.

A.2 СФЕРА ЗАСТОСУВАННЯ, ТЕРМІНИ ТА ЇХ ВИЗНАЧЕННЯ (APPLICABILITY AND DEFINITIONS)

Дана Ліцензія застосовується до будь-якого посібника чи іншої роботи, яка містить повідомлення, розміщене правовласником, згідно якого остання може розповсюджуватись на умовах даної Ліцензії (далі ліцензійне повідомлення). Таке повідомлення надає всесвітню, необмежену в часі, безгонорарну ліцензію на використання даної роботи на умовах зазначених тут. Термін "Документ" далі в тексті означає будь-який подібний посібник чи працю. Будь яка особа є ліцензіатом і далі адресується як "Ви". Ви погоджуєтеся з умовами ліцензії, якщо Ви копіюєте, вносите зміни чи розповсюджуєте роботу у спосіб, який вимагає на це отримання дозволу згідно законодавства про авторське право.

"Модифікована версія" Документу — це будь-яка робота, яка містить Документ або його частину, скопійовані дослівно чи зі змінами і/або перекладені на іншу мову.

"Другорядний розділ"— це додаток, що має назву, або передмова до Документу, в якому розглядаються виключно питання стосунку (зв'язку, взаємовідношень, ставлення) видавців чи авторів Документу до змісту Документу в цілому (або до питань, пов'язаних зі змістом Документу) та не міститься нічого, що є частиною цього змісту. (Так наприклад, якщо Документ частково є

посібником з математики, в Другорядному розділі не може міститись нічого, що має відношення безпосередньо до математики). Під ставленням маються на увазі питання історичних зв'язків зі змістом Документу чи спорідненою темою, або правові, комерційні, філософські, етичні чи політичні погляди стосовно останніх.

"Незмінні розділи"— це певні Другорядні розділи, заголовки яких перераховано в ліцензійному повідомленні. Якщо розділ не підпадає під наведене вище визначення Другорядного, він не може бути визначений як Незмінний. Документ може не мати Незмінних розділів. Якщо в Документі не вказано жодного Незмінного розділу, то такі відсутні.

"Текст обкладинки"— це певний короткий уривок тексту, який використано як "Текст титульної обкладинки" або "Текст зворотньої обкладинки". Текст титульної обкладинки може містити максимум 5 слів, а Текст зворотньої обкладинки — максимум 25 слів.

"Прозорий" екземпляр Документу — це машино-читабельна копія Документу, представлена в форматі, специфікація якого доступна для загального використання, і який придатний для перегляду безпосередньо за допомогою загальнодоступних текстових редакторів, або (для зображень складених з пікселів) загальнодоступних редакторів растрової графіки, або (для креслень) деяких широко доступних редакторів векторної графіки. Вказаний формат повинен бути придатний для вводу в програму форматування текстів або автоматичного конвертування в один з форматів, який, в свою чергу, придатний для вводу в програму форматування текстів. Екземпляр документу, представлений в форматі, розмітка якого або її відсутність перешкоджає чи не дозволяє вносити зміни наступному користувачу, не є Прозорим. Формат зображень не є Прозорим, якщо використовується для значної частини тексту. Екземпляр Документу, який не є Прозорим, називається "Непрозорим".

Прикладами форматів, в яких може бути представлений Прозорий екземпляр Документу, є простий формат ASCII без розмітки, формат вводу Texinfo, формат вводу LaTeX, SGML або XML з використанням загальнодоступного DTD, відповідний стандартам простий HTML, PostScript або PDF формати, які можуть редагуватися людиною. Прикладами Прозорих форматів зображень є PNG, XCF та JPG. Непрозорі формати включають патентовані формати, які можна прочитати і модифікувати тільки за допомогою запатентованих текстових процесорів, права на використання яких вільно не передаються, SGML або XML для яких DTD та/або інструменти обробки не є загальнодоступними, згенерований машиною HTML, PostScript або PDF, які є результатом роботи деяких текстових процесорів і призначені тільки для виводу.

"Титульна сторінка" для друкованої книги означає титульну сторінку саму по собі і наступні за нею сторінки, які необхідні для розміщення на них, розбірливо, матеріалу, який згідно даної Ліцензії повинен бути розміщений на титульній сторінці. Для роботи з форматами, які не передбачають наявності титульної сторінки як такої, "Титульна сторінка" означає текст, який розміщено після назви праці, надрукованої найбільш помітним шрифтом, та перед початком основного тексту.

"Озаглавлений розділ XYZ"— це частина Документу, виділена в окремий розділ, який має заголовок "XYZ" або ж, у випадку перекладу заголовку на іншу мову, "XYZ" є частиною заголовку і міститься в круглих дужках після тексту, яким його перекладено. (Під "XYZ" мається на увазі безпосередньо назва розділу, наприклад, "Подяка", "Присвячення", "Погодження" або "Історія"). "Зберегти заголовки" такого розділу в зміненому Вами Документі означає залишити його "Озаглавленим розділом" "XYZ" відповідно до вищенаведеного визначення.

Документ може містити пункт "Гарантійні Застереження" відразу після ліцензійного повідомлення. Такі Гарантійні Застереження вважаються частиною даної Ліцензії, але лише як відмова в наданні гарантій: будь-який інший під-

текст Гарантійних Застережень не має юридичної сили і не впливає на трактування даної Ліцензії.

A.3 КОПІЮВАННЯ БЕЗ ЗМІН (VERBATIM COPYING)

Ви можете копіювати і розповсюджувати Документ в будь-який спосіб, на комерційній чи некомерційній основі, при умові, що кожен екземпляр Документу містить текст даної Ліцензії, повідомлення про авторське право та повідомлення, в якому зазначено, що Документ доступний на умовах даної Ліцензії, та Ви не додали жодних інших умов до вже зазначених в Ліцензії. У створених Вами копіях Ви не маєте права використовувати технічні засоби, які б контролювали читання або перешкоджали читанню Документа та його подальшому копіюванню та розповсюдженню. Ви маєте право приймати компенсацію за створені Вами копії. Якщо ви розповсюджуєте достатньо велику кількість копій, Ви повинні також дотримуватися умов, зазначених в розділі 3.

Ви маєте право позичати та демонструвати (робити доступними для загального користування) створені Вами копії на умовах зазначених вище.

A.4 ТИРАЖУВАННЯ (COPYING IN QUANTITY)

Якщо Ви видаєте друковані копії Документу в кількості більше 100 штук і в ліцензійному повідомленні зазначено Текст обкладинки, Ви повинні використати його для обкладинок у створених Вами копіях: як Текст титульної обкладинки так і Текст зворотньої обкладинки. На обох сторонах обкладинки повинно бути чітко і розбірливо зазначене Ваше ім'я, як видавника копій. На титульній обкладинці повинна бути вказана повна назва документу без пропусків і скорочень і всі слова з назви повинні бути надруковані чітко і розбірливо. Ви маєте право розмістити на обкладинці додаткові відомості. Копіювання зі змінами, які стосуються тільки обкладинки, при умові збереження назви Документу та дотриманні умов, зазначених вище, вважаються копіюванням без змін.

Якщо об'єм тексту, який повинен бути розміщений на будь-якій з обкладинок, занадто великий, щоб бути надрукованим розбірливо, Ви можете розмістити його частину на обкладинці (так багато, як вважаєте прийнятним), а решту на прилеглих сторінках.

Якщо Ви видаєте чи розповсюджуєте Непрозорі копії Документу в кількості більше 100 штук, Ви повинні додати до кожної такої копії машино-читабельну Прозору копію, або зазначити загальнодоступну мережеву адресу, з якої, використовуючи стандартні мережеві протоколи, можна завантажити повну Прозору копію Документу без доданих в Непрозорій копії змін. Якщо Ви вирішили скористатися останньою опцією, Ви повинні вжити необхідних кроків для забезпечення доступу до Прозорої копії за вказаною мережевою адресою якнайменше на протязі року після дати останнього розповсюдження Непрозорої копії даного видання (особисто чи через агентів або торгових посередників).

Перед розповсюдженням великої кількості копій рекомендується, але не вимагається, повідомити про це авторів Документу, щоб дати їм можливість надіслати Вам останню версію Документу.

A.5 МОДИФІКАЦІЇ (MODIFICATIONS)

Ви можете копіювати і розповсюджувати Модифіковану версію Документу за дотримання вимог, зазначених вище в розділах 2 та 3, та при умові, що Модифікована версія публікується у відповідності до даної Ліцензії, тобто Ви передаєте

права на розповсюдження і внесення змін до Модифікованої версії Документу аналогічні правам на розповсюдження і внесення змін до Документу. При копіюванні і розповсюдженні Модифікованої версії Документу Ви зобов'язані:

- А. Помістити на Титульній сторінці (та обкладинках, якщо такі є) заголовок, відмінний від заголовку Документу та всіх його попередніх версій (які, якщо такі є, повинні бути перелічені в розділі "Історія"). Ви можете використати такий же заголовок, як і для попередньої версії Документу, лише за умови надання згоди на це видавця попередньої версії.
- В. Вказати на Титульній сторінці осіб, які є авторами змін до Модифікованої версії Документу, та, щонайменше, п'ять головних авторів Документу (всіх головних авторів, якщо їх менше п'яти), якщо вони не звільнили Вас від даної вимоги.
- С. Вказати на Титульній сторінці ім'я видавця Модифікованої версії та зазначити, що дана особа є видавцем.
- D. Зберегти в Документі всі повідомлення про авторське право.
- Е. Додати відповідне повідомлення про авторське право на внесені Вами зміни поряд з іншими існуючими повідомленнями про авторське право.
- F. Помістити, відразу після повідомлення про авторське право, повідомлення, яке дає право кожному використовувати Модифіковану версію Документу на умовах даної Ліцензії (ліцензійне повідомлення), в форматі, наведеному нижче в Додатку до даної Ліцензії.
- G. Зберегти в ліцензійному повідомленні для Модифікованої версії повний список Незмінних розділів та Тексти обкладинок, зазначені в ліцензійному повідомленні Документу.
- Н. Включити в Модифіковану версію Документу повний текст даної Ліцензії.
- I. Зберегти Озаглавлений розділ "Історія": залишити незмінним його заголовок та додати в нього пункт, де вказати, щонайменше, заголовок, рік, нових авторів та видавця Модифікованої версії Документу (як на Титульній сторінці). Якщо в Документі відсутній Озаглавлений розділ "Історія", Ви зобов'язані створити такий розділ в Модифікованій версії, помістивши в нього заголовок, рік, авторів та видавця Документу (як на Титульній сторінці) та додавши пункт, який стосується Модифікованої версії і описаний вище.
- J. Вжити необхідних кроків для збереження мережевої адреси, вказаної в Документі (якщо така є), за якою можна отримати доступ до Прозорої копії Документу, та мережевих адрес, за якими знаходяться копії попередніх версій Документу. Такі адреси можна розмістити в розділі "Історія". Ви можете опустити мережеві адреси для версій, які були опубліковані щонайменше за чотири роки до публікації Документу, або видавець яких дав на це згоду.
- K. В Озаглавлених розділах "Подяка" або "Присвячення" залишити незмінним заголовок, а також зміст і стиль всіх подяк та/чи присвячень.
- L. Зберегти всі Незмінні розділи Документу, залишивши незмінними їх текст та заголовки. Послідовний номер розділу, або його еквівалент, не вважаються частиною заголовку.
- M. Видалити Озаглавлений розділ "Погодження", якщо такий існує. Даний розділ не може бути включений в Модифіковану версію.

- N. Не присвоювати жодному з існуючих розділів заголовки "Погодження" або заголовки, який повторює назву одного з Незмінних розділів.
- O. Зберегти існуючі Гарантійні Застереження.

Якщо Модифікована версія містить нові розділи передмови або додатки, які класифікуються як Другорядні розділи і не містять матеріалів, скопійованих з Документу, Ви можете, на свій розсуд, визначити всі або окремі такі розділи як незмінні. Для цього необхідно додати їх назву до списку Незмінних розділів в ліцензійному повідомленні Модифікованої версії. Заголовки таких розділів повинні відрізнятися від вже існуючих.

Ви можете додати Озаглавлений розділ "Погодження", якщо він містить тільки погодження Модифікованої версії різними сторонами. Наприклад, констатацію рецензування компетентними особами, або посилання на те, що текст було затверджено організацією як офіційне визначення стандарту.

В Модифікованій версії Ви можете додати до п'яти слів до Тексту титульної обкладинки і до двадцяти п'яти слів до Тексту зворотної обкладинки в кінець списку "Текст обкладинки" ліцензійного повідомлення. При цьому, кожною особою, безпосередньо або від її імені, може бути додано лише одну фразу до Тексту титульної обкладинки і одну — до Тексту зворотної обкладинки. Якщо Документ вже має текст обкладинки, доданий Вами, чи іншою стороною, яка діяла від Вашого імені, Ви не можете додати новий текст, але можете замінити старий при умові отримання на це чітко висловленої згоди попереднього видавця, який додав старий текст.

Автор(и) та видавець(ці) Документу за умовами даної Ліцензії не передають право використовувати їх імена в цілях реклами чи з метою створення враження про їх згоду з будь-якою Модифікованою версією.

A.6 ОБ'ЄДНАННЯ ДОКУМЕНТІВ] (COMBINING DOCUMENTS)

Ви можете об'єднати Документ з іншими документами, випущеними на умовах даної Ліцензії, при дотриманні умов, перерахованих в розділі 4. При цьому, Ви повинні включити в об'єднаний документ всі Незмінні розділи з оригінальних документів і перерахувати їх як Незмінні розділи в ліцензійному повідомленні об'єднаного документу, а також зберегти Гарантійні Застереження.

Об'єднаний документ повинен містити лише одну копію тексту даної Ліцензії, а повторювані Незмінні розділи можуть бути замінені єдиною копією. Якщо повторювані Незмінні розділи мають однакові заголовки, але відмінний зміст, до їх заголовків необхідно додати, в круглих дужках, ім'я їх автора або видавця, якщо такий відомий, або унікальний номер. Відповідні зміни повинні бути внесені до заголовків Незмінних розділів в ліцензійному повідомленні об'єднаного документу.

В новому Документі Ви повинні об'єднати всі Озаглавлені розділи "Історія" з оригінальних документів в один Озаглавлений розділ "Історія"; подібним чином об'єднати Озаглавлені розділи "Подяка" та "Присвячення" та видалити всі Озаглавлені розділи "Погодження".

A.7 ЗБІРНИКИ ДОКУМЕНТІВ (COLLECTIONS OF DOCUMENTS)

Ви маєте право створити збірник з Документу та інших документів, які розповсюджуються на умовах даної Ліцензії, і включити в нього лише одну копію даної Ліцензії при умові дотримання вимог, зазначених в ній, щодо дослівного копіювання кожного документу.

Ви маєте право виділити окремий документ з такого збірника і розповсюджувати його окремо, в відповідності з даною Ліцензією та при умові, що текст

даної Ліцензії включено в виділений документ і дотримані всі інші умови даної Ліцензії щодо дослівного копіювання документу.

A.8 ПОЄДНАННЯ З НЕЗАЛЕЖНИМИ РОБОТАМИ (AGGREGATION WITH INDEPENDENT WORKS)

Поєднання Документу, або його похідних, з іншими окремими і самостійними документами чи працями на одному носії інформації чи пристрої для її розповсюдження називається "збіркою", при умові, що авторські права на неї не обмежують прав користувачів, які ті мають на окремі документи в збірці. Коли Документ входить до збірки, дана Ліцензія не поширюється на інші праці в збірці, якщо ті не є похідними від Документу.

Якщо для копій Документу діють вимоги щодо Текстів обкладинок (розділ 3), то, у випадку, коли Документ складає менше половини збірки, Тексти обкладинок може бути розміщено на обкладинках, в які Документ заключено всередині збірки, або електронному еквіваленті таких обкладинок, якщо Документ знаходиться в електронній формі. Інакше, Тексти обкладинок повинні бути надруковані на обкладинках, в які заключено всю збірку.

A.9 ПЕРЕКЛАД (TRANSLATION)

Переклад вважається однією з різновидностей модифікації Документу і перекладені версії Ви можете розповсюджувати на умовах, перерахованих в розділі 4. Заміна Незмінних розділів перекладеною версією вимагає отримання особливого дозволу від їх авторів, але перекладені Незмінні розділи можуть бути включені на додаток до збережених (не перекладених) Незмінних розділів. Ви маєте право включити в перекладений документ переклад даної Ліцензії, ліцензійного повідомлення та Гарантійних Застережень при умові, що Ви включили також оригінальну версію Ліцензії англійською мовою та оригінальні версії ліцензійного повідомлення та застереження. У випадку суперечностей між оригінальною версією даної Ліцензії та її перекладом оригінальна версія має перевагу.

Вимога даної Ліцензії зберегти заголовки Озаглавлених розділів "Подяка", "Присвячення", або "Історія" Документу (розділи 1 та 4 даної Ліцензії) у випадку перекладу типово вимагає зміни оригінального заголовку.

A.10 ПРИПИНЕННЯ (TERMINATION)

Ви не маєте права копіювати, модифікувати, субліцензувати чи розповсюджувати Документ в жоден спосіб, окрім зазначеного умовами даної Ліцензії. Будь-які інші способи копіювати, модифікувати, субліцензувати чи розповсюджувати Документ не мають юридичної сили і тягнуть за собою припинення Ваших прав згідно даної Ліцензії. Однак, права сторін, які отримали від Вас на умовах даної Ліцензії копії Документу чи права на нього, залишаються в силі при умові повного дотримання ними вимог даної Ліцензії.

A.11 МАЙБУТНІ ПЕРЕГЛЯДИ ЦЬОЇ ЛІЦЕНЗІЇ (FUTURE REVISIONS OF THIS LICENSE)

Free Software Foundation має право публікувати нові, переглянуті версії GNU Free Documentation License. В таких версіях буде збережено всі основні принципи, закріплені в даній Ліцензії, але може бути додано додаткові пункти, які регулюють правовідносини, які виникли після публікації попередніх версій Ліцензії. Див. <http://www.gnu.org/copyleft/>.

Кожній Версії Ліцензії присвоюється свій власний номер. Якщо в Документі вказано певну версію Ліцензії та зазначено "або будь-яка наступна версія", Ви маєте право посилалися на терміни та умови або зазначеної версії, або будь-якої наступної версії, опублікованої (не як проект) Free Software Foundation. Якщо в Документі не вказано версію Ліцензії, Ви маєте право посилалися на будь-яку версію Ліцензії, будь-коли опубліковану (не як проект) Free Software Foundation.

A.12 ДОДАТОК

Як використовувати дану Ліцензію для Ваших документів

Для використання даної Ліцензії в створеному Вами документі повний текст даної Ліцензії необхідно включити в документ та розмістити відразу після титульної сторінки документу ліцензійне повідомлення наступного змісту:

Copyright © РІК ПУБЛІКАЦІЇ ІМ'Я АВТОРА.

Кожен має право копіювати, розповсюджувати та вносити зміни до даного Документу на умовах GNU Free Documentation License, Версії 1.2 або будь-якої наступної версії, опублікованої Free Software Foundation;

даний Документ не містить Незмінних розділів;

даний Документ не має Текстів титульної та зворотньої обкладинок.

Копія даної ліцензії включена в розділ з заголовком "GNU Free Documentation License".

Якщо Ваш документ має Незмінні розділи та Текст обкладинок, вони повинні бути зазначені в ліцензійному повідомленні:

даний Документ має Незмінні розділи

ЗАЗНАЧИТИ ЗАГОЛОВКИ ТАКИХ РОЗДІЛІВ та

Текст титульної обкладинки **ЗАЗНАЧИТИ** та

Текст зворотньої обкладинки **ЗАЗНАЧИТИ**

Якщо Ваш документ має Незмінні розділи, але не має Тексту обкладинок, тощо, скористайтесь вищенаведеними прикладами для компоновання ліцензійного повідомлення, яке б відповідало такій комбінації.

Якщо Ваш документ містить нетривіальні приклади програмного коду, ми рекомендуємо Вам опублікувати його паралельно на умовах відкритої ліцензії на програмне забезпечення, на Ваш вибір, наприклад, GNU General Public License, щоб дозволити його використання в відкритому програмному забезпеченні.

Зміст

1	Мова запитів	3
1.1	Запити про сегменти	3
1.2	Запити про початкові форми	6
1.3	Запити вищого порядку	6
1.4	Запити про морфосинтаксичні теги	8
1.5	Багатозначності і уоднозначення	10
1.6	Обмеження запиту до речення або абзацу	12
1.7	Обмеження запиту з допомогою метаданих	13
1.8	Вирівнювання результатів	14
2	Poliqarp	15
2.1	Інтернет-версія	16
2.2	Графічна версія	19
2.3	Текстова версія	22
2.3.1	Редагування запитів у командному рядку	25
2.3.2	Конфігурація програми	25
2.3.3	Системні змінні	30
A	GNU Free Documentation License	
	Версія 1.2	
	Листопад 2002	31
A.1	ПРЕАМБУЛА (PREAMBLE)	31
A.2	СФЕРА ЗАСТОСУВАННЯ, ТЕРМІНИ ТА ЇХ ВИЗНАЧЕННЯ (APPLICABILITY AND DEFINITIONS)	31
A.3	КОПІЮВАННЯ БЕЗ ЗМІН (VERBATIM COPYING)	33
A.4	ТИРАЖУВАННЯ (COPYING IN QUANTITY)	33
A.5	МОДИФІКАЦІЇ (MODIFICATIONS)	33
A.6	ОБ'ЄДНАННЯ ДОКУМЕНТІВ (COMBINING DOCUMENTS)	35
A.7	ЗБІРНИКИ ДОКУМЕНТІВ (COLLECTIONS OF DOCUMENTS)	35
A.8	ПОЄДНАННЯ З НЕЗАЛЕЖНИМИ РОБОТАМИ (AGGREGATION WITH INDEPENDENT WORKS)	36
A.9	ПЕРЕКЛАД (TRANSLATION)	36
A.10	ПРИПИНЕННЯ (TERMINATION)	36
A.11	МАЙБУТНІ ПЕРЕГЛЯДИ ЦЬОЇ ЛІЦЕНЗІЇ (FUTURE REVISIONS OF THIS LICENSE)	36
A.12	ДОДАТОК Як використовувати дану Ліцензію для Ваших документів	37