

УДК 004.421

Т.М. Заболотня, А.Ю. Михайлюк,
В.П. Тарасенко

СТРУКТУРНО-АЛГОРИТМІЧНА ОРГАНІЗАЦІЯ ПРОГРАМНОГО АГЕНТА-ОРФОКОРЕКТОРА

Вступ

На сьогодні до числа найбільш затребуваних класів програмного забезпечення належать документоорієнтовані інформаційні системи (ДІС). Для підтримки конкурентоспроможності таких систем на світовому ринку програмного забезпечення (ПЗ) їх розробляють як відкритий програмний продукт, придатний до масштабування. Останнім часом в основу реалізації такого типу програм все частіше буває покладено агентоорієнтований підхід [1], відповідно до якого модулі, що входять або мають увійти до складу відкритих інформаційних систем, повинні будуватися як програмні агенти [1–3].

У даній статті розглядається проблема побудови одного з найпоширеніших інструментів для обробки текстових даних – орфокоректора. Зараз він найчастіше розробляється або як жорстко вбудований в ДІС модуль, що здійснює виправлення помилок, або як автономна програма орфокорекції, до якої при необхідності звертається сам користувач. Обидва наведені способи організації коректорів, поперше, не передбачають можливості його підключення як окремо розробленої утиліти до програмних комплексів із відкритою архітектурою [2], а по-друге, зумовлюють необхідність побудови великої кількості модифікацій подібних модулів.

Постановка задачі

Для спрощення інкапсуляції та ефективного функціонування орфокоректора в ДІС необхідно в основу його розробки покласти агентний підхід. Звідси метою статті є визначення структурно-алгоритмічної організації агента-коректора та його поведінки в агентному середовищі.

Документоорієнтована система як зовнішнє середовище агента-коректора

Особливою рисою програмних агентів є необхідність їх функціонування в деякому

зовнішньому середовищі, з яким вони можуть взаємодіяти, але яке не мають можливості контролювати. Як правило, вибір конкретної дії агент здійснює, базуючись на даних про поточний стан зовнішнього середовища. В загальному вигляді агент можна подати як функцію [1]

$$\text{action} : S \rightarrow A, \quad (1)$$

де S – множина станів середовища; A – множина дій, які здатен виконувати агент.

Для агента-коректора зовнішнім середовищем є документоорієнтована інформаційна система, до складу якої він входить, та програмно-апаратне забезпечення цієї системи. Стан даного середовища пропонується визначати за набором значень квазістатичних параметрів ДІС та оперативними даними, які надходять до агента-коректора у вигляді повідомлень від інших компонентів ДІС або від користувача.

Відзначимо, що значення квазістатичних параметрів ДІС та оперативні дані впливають не тільки на вибір алгоритму роботи коректора в реальному часі, але і на структурно-алгоритмічну організацію останнього, оскільки нею має бути передбачена можливість роботи агента-коректора за різних умов.

Вибір алгоритму роботи агента-коректора виключно на основі аналізу стану зовнішнього середовища не завжди гарантує його ефективне функціонування в межах багатокомпонентної відкритої ДІС. Для забезпечення більш гнучкої та надійної роботи орфокоректора пропонується будувати його як агента, дії якого визначаються також його внутрішнім станом. У загальному вигляді такий агент-коректор подамо як функцію

$$\text{action} : I \times S \rightarrow A, \quad (2)$$

де I – множина внутрішніх станів орфокоректора, які він змінює під впливом стану зовнішнього середовища. Пропонується до множини I внести дані про поточний стан складових агента-коректора, а також дані про можливість ініціювання роботи орфокоректора іншим модулем ДІС.

Подальше ускладнення функції action (а саме, проектування агента-коректора здатним до створення цілей, до прогнозування зміни власного стану та стану зовнішнього середовища, наділення його ментальними властивостями [1] тощо) вважаємо зайвим, оскільки орфокоректор є сервісним доповненням ДІС і до його специфікації не повинно входити

ініціювання взаємодії з іншими модулями ДІС. Крім того, занадто ускладнювати організацію агента-коректора (особливо при локальному розміщенні ДІС) не слід, через те що це може спричинити втрату оперативності його функціонування, що для програмного забезпечення допоміжного характеру вкрай небажане. Таким чином, по відношенню до ДІС агент-коректор має бути реактивним.

Розглянемо квазістатичні параметри і оперативні дані документоорієнтованих інформаційних систем.

Квазістатичні параметри ДІС. Значення цих параметрів під час роботи ДІС змінюються дуже рідко (якщо змінюються взагалі). До них насамперед належать:

а) структурна організація (будова) ДІС: локальне чи віддалене розміщення її компонентів, кількість модулів, які можуть ініціювати роботу агента-коректора, наявність серед програмних ресурсів, які використовує ДІС, таких, що можна залучати до процесу визначення варіантів виправлення слова з орфографічною помилкою (далі – *спотворене слово*);

б) функціональні характеристики ДІС:

- ступінь складності (інтелектуальності) алгоритмів аналізу текстових даних ДІС – визначає загальну необхідність проведення синтактико-семантичного узгодження варіантів виправлення спотвореного слова з його контекстним оточенням чи будь-якими іншими даними;

- характер текстових даних, з якими працює ДІС: формалізований текст, текст, в якому переважають числові дані, структурований текст, художній текст, спеціалізована термінологія тощо;

- пріоритетність критеріїв визначення ефективності роботи ДІС: надання переваги максимальному ступеню автоматизації процесів обробки даних, точності отримуваних результатів, швидкості виконання алгоритмів тощо;

в) програмно-апаратне забезпечення роботи ДІС (і агента-коректора, зокрема): обчислювальна потужність апаратних ресурсів, пропускна здатність каналу передачі даних у випадку віддаленого розміщення модулів ДІС, параметри операційної системи, під керуванням якої працює ДІС, можливості СУБД, які використовуються для збереження даних, тощо.

Звісно, існують параметри, які характеризують ДІС, але які не були згадані вище (наприклад, детальні дані про склад ДІС, про функціонування кожного її модуля тощо). Їх не

було внесено до розгляду через те, що значення цих параметрів ніяк не впливають на архітектуру та поведінку агента-коректора. Тому можна сказати, що згадані квазістатичні параметри та оперативні дані визначають групи станів зовнішнього середовища, які є еквівалентними з точки зору їх сприйняття агентом-коректором.

Проаналізуємо вплив кожної складової множин S та I на структурно-алгоритмічну організацію агента-коректора.

Структурна організація ДІС. На рис. 1 наведені приклади дислокації агента-коректора у ДІС за різних способів організації останньої.

У випадку *локального* розміщення компонентів ДІС орфокоректор доцільно реалізувати у формі одного *реактивного* агента (рис. 1,а) [3]. При цьому компактність ДІС дозволяє коректору не мати власних лінгвістичних програмних ресурсів, якщо їх необхідний набір входить до складу ДІС (на рис. 1 можливий зв'язок між коректором та ресурсами ДІС зображений штриховою лінією).

Побудова коректора у вигляді сукупності реактивних агентів Ag_t ; (наприклад, для забезпечення паралельного виконання певних незалежних етапів алгоритму виправлення помилок [4]) має значення тільки для ДІС, що функціонує на багатопроесорному комп'ютері.

Форма реалізації орфокоректора в умовах *віддаленого* розміщення складових ДІС залежить від взаємного розміщення модулів (M_i) та ресурсів (Res_i), з якими він взаємодіє в процесі визначення варіантів виправлення спотвореного слова. Чим більше компонентів, необхідних для роботи коректора, знаходиться на одному з ним вузлі комп'ютерної мережі, тим швидше будуть отримані результати корекції і тим менше даних передаватиметься при цьому каналом зв'язку (рис. 1,б). Випадок, коли всі потрібні компоненти належать одному вузлу, можна розглядати як локальне розміщення ДІС та коректора. Тоді, відповідно, коректор можна реалізувати у вигляді єдиного реактивного агента.

Якщо ж алгоритмом виправлення помилок передбачено виконання складної послідовності операцій із використанням великої кількості віддалених лінгвістичних програмних ресурсів (наприклад, при багаторівневому аналізі текстових даних), то допускається організація орфокоректора у формі *багатоагентної* системи, кожна складова якої буде відповідати за організацію роботи з окремим ресурсом або реалізо-

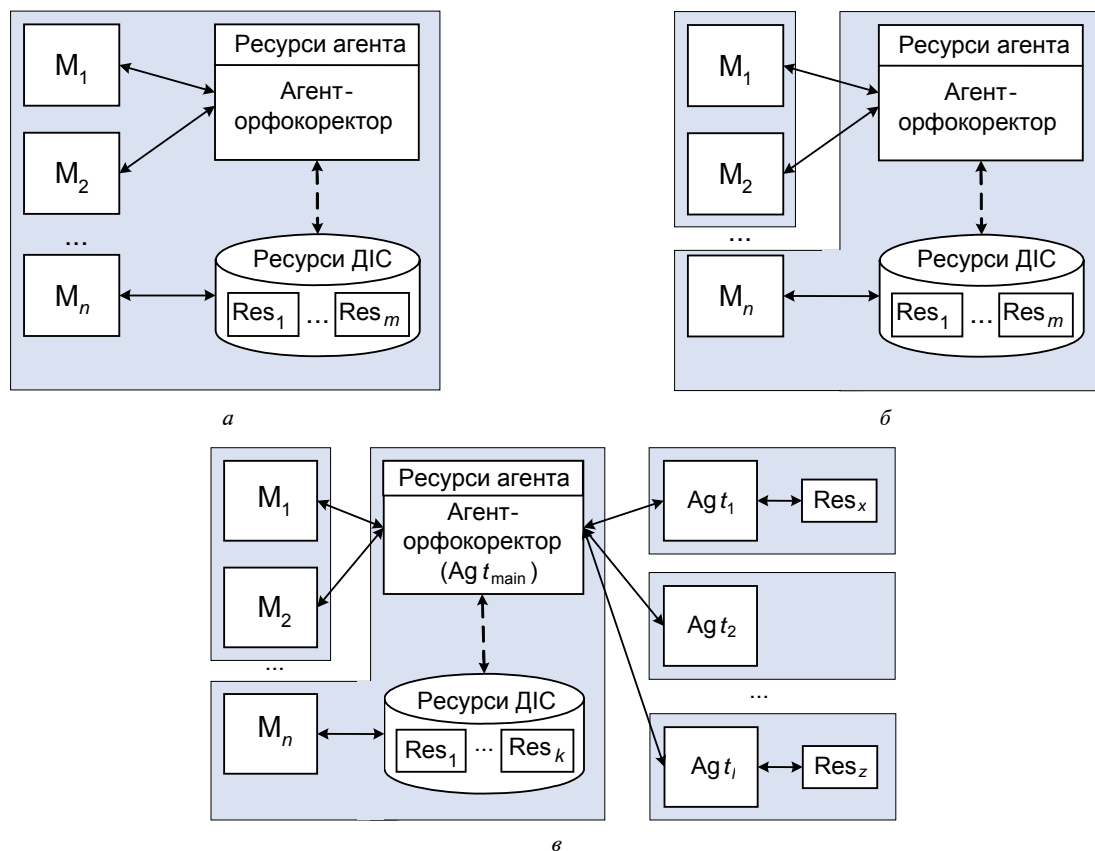


Рис. 1. Варіанти реалізації агента-коректора залежно від розміщення компонентів ДІС

увати певний етап процесу корекції помилок у спотвореному слові (рис. 1, в). При цьому агент $Ag t_{main}$, який безпосередньо взаємодіє з ДІС, з одного боку, має контролювати роботу всіх інших агентів $Ag t_i$, а з іншого, залишатися водночас реактивним відносно компонентів ДІС.

Також потрібно зазначити, що звернення коректора до ресурсів ДІС може спричинити зниження її загальної надійності та швидкодії.

Тому при розробці структурно-алгоритмічної організації кожного орфокоректора необхідно проаналізувати доцільність внесення до його складу реплік деяких ресурсів ДІС (звичайно, за умови, якщо відповідне технічне забезпечення дозволяє це зробити). Якщо ж ДІС не має у своєму складі потрібних для роботи коректора ресурсів, то їх необхідно внести або до структури самого агента, або до складу ДІС (з можливістю їх використання й іншими модулями).

Порівняно з локальним варіантом розміщення ДІС при віддаленому розташуванні її компонентів зростає ймовірність їх одночасного звернення до коректора. Через це для будь-

якої структури ДІС необхідно передбачити механізм попередження конфліктів доступу до коректора. Так, або спеціальний супервізорний модуль ДІС може здійснювати моніторинг зайнятості агента-коректора, або орфокоректор має самостійно контролювати спроби одночасного звернення до нього з боку кількох модулів ДІС.

Функціональні характеристики ДІС. Кожна документоорієнтована система має певні особливості функціонування. У цьому розумінні такі системи відрізняються між собою даними, якими їм доводиться оперувати, алгоритмами роботи та критеріями оцінювання ефективності останніх. Для того щоб агент-коректор міг гнучко налаштовуватися на роботу в складі ДІС різних функціональних типів, необхідно передбачити в його алгоритмічній організації набір відповідних сценаріїв поведінки, які будуть вибиратися залежно не тільки від будови ДІС, але й від функціональних характеристик ДІС.

Ступінь складності алгоритмів обробки текстових даних. На сьогодні існуючі документоорієнтовані інформаційні системи залежно від свого призначення реалізують різні за складністю алгоритми аналізу текстових даних. Мож-

на стверджувати, що в переважній більшості випадків ступінь складності алгоритмів обробки спотвореного слова відповідним орфокоректором залежить від рівня аналізу текстів, який має місце в ДІС в цілому. Так, для програми, яка оперує даними на рівні морфологічного аналізу тексту (наприклад, для простого редактора тексту), непотрібно будувати коректор, що буде узгоджувати варіанти виправлення із змістом контекстного оточення слова, в якому допущена помилка. Для великих же аналітичних комплексів цілком реальною є необхідність у детальному синтактико-семантичному аналізі спотвореного слова.

Відповідно до рівня складності алгоритмів роботи орфокоректора визначається і набір ресурсів, які він використовує: від невеликих словників, що містять короткий список лексем, до масштабних семантичних ресурсів та онтологій предметної галузі.

Характер текстових даних. Від характеру текстових даних, якими оперує ДІС, залежить вибір способу їх обробки та ресурсів, які при цьому будуть використовуватися як самою системою, так і орфокоректором.

Наприклад, для виправлення помилки у словах, що належать до термінології певної предметної галузі, необхідним є залучення теоретично й емпірично отриманих закономірностей будови спеціалізованої лексики та її вживання в документах, а також предметноорієнтованих словникових ресурсів, які дозволили б адекватно ідентифікувати та виправити спотворене слово. Корекція формалізованого тексту (програмного коду або документа, складеного будь-якою штучною мовою) базується на правилах складання текстів такого характеру та на залученні до процесу виправлення ресурсів, які містять перелік мовних конструкцій, притаманних даним формалізованій мові. Для знаходження вірного варіанта написання спотвореного слова з художнього тексту необхідне проведення детального синтактико-семантичного аналізу контексту даного слова з використанням словників, які містять інформацію про слова, що можуть мати переносне значення, усталені мовні звороти та інші прийоми художньої виразності. Якщо ж текст, в якому здійснюється орфокорекція, насичений числовими даними, то застосування алгоритмів та ресурсів, які використовуються для виправлення помилок у словах, не є доцільним. Замість цього потрібно використовувати способи зна-

ходження спотворень у послідовностях цифр, які мають місце в галузі кодування.

Пріоритетність критеріїв оцінки ефективності роботи ДІС. Вибір алгоритму роботи агента-коректора залежить і від критерію, який використовується для оцінки ефективності функціонування ДІС в цілому. Якщо ДІС вважається ефективною в разі забезпечення нею високого ступеня автоматизації її роботи, то орфокоректор як складова такої системи також має застосовувати алгоритми, що передбачають мінімальне втручання користувача. У випадку, коли ДІС працює в режимі, який потребує оперативної обробки даних, агент-коректор повинен вибирати для реалізації процесу корекції алгоритми, які характеризуються високою швидкістю роботи. Аналогічно пріоритетність точності обробки текстових даних у ДІС означає використання відповідних алгоритмів роботи орфокоректора.

Програмно-апаратне забезпечення роботи ДІС. Від характеристик апаратних ресурсів та програмного забезпечення, на базі яких функціонує ДІС, залежить допустима складність алгоритмів обробки текстових даних (у тому числі орфокорекції), обсяг та структурна організація даних у використовуваних ресурсах, принципова можливість віддаленого розміщення компонентів ДІС, певні аспекти, пов'язані із комунікацією агента-коректора із зовнішнім середовищем та з його роботою під керуванням різних операційних систем тощо.

Наприклад, при організації роботи агента-коректора із словниковими ресурсами (які, зазвичай, організовані у формі БД) необхідно враховувати особливості СУБД, на базі якої вони реалізовані, і будувати для кожного випадку свою множину запитів до БД та алгоритми обробки результатів виконання цих запитів. Певні ж характеристики каналу зв'язку між орфокоректором та модулями ДІС можуть накладати обмеження на розмір, тип кодування та принцип адресації повідомлень, якими програмні компоненти обмінюються один з одним.

Таким чином, нами розглянуто вплив значень квазістатичних параметрів ДІС на структурно-алгоритмічну організацію агента-коректора як складової цієї системи. Оскільки ці значення змінюються дуже рідко, то не має сенсу передавати їх до орфокоректора щоразу при необхідності ініціювати процес виправлення помилок у спотвореному слові. Доцільно заносити відомості про поточну комбінацію

значень наведених параметрів до внутрішніх інформаційних масивів агента на етапі його ініціалізації в ДІС та оновлювати відповідно до змін у системі.

Оперативні повідомлення від ДІС. Оперативні дані (повідомлення), які надходять до сенсорів агента-коректора від інших компонентів ДІС або від користувача під час роботи в реальному часі, можна розділити на два види: службові повідомлення та запити на виправлення спотвореного слова з набором вказівок щодо особливостей проведення процесу корекції.

Для того щоб ініціювати роботу орфоко-ректора, передавати йому необхідні текстові дані, а також додаткові вказівки (інструкції) щодо процесу виправлення помилок, відповідний модуль ДІС чи користувач мають сформу-вати певне оперативне повідомлення і наді-слати його до орфоко-ректора. На основі такого повідомлення, а також виходячи з відомостей про значення квазістатичних параметрів, агент-коректор динамічно визначає оптимальний для поточної ситуації алгоритм корекції, а також набір ресурсів, які будуть використовуватися під час роботи цього алгоритму.

Переважну більшість оперативних повідом-лень становлять безпосередні запити на ви-правлення спотвореного слова з набором допо-міжних інструкцій щодо проведення процесу корекції. Кожен запит умовно поділяється на дві частини: основні дані та додаткові інструк-ції. До основних даних належить спотворене слово та його контекст (у разі, коли потрібно здійснювати синтактико-семантичний аналіз контекстного оточення). Інструкції можуть мі-стити дані щодо:

- максимальної кількості помилок, яку має виправляти коректор (чим менше поми-лок, тим швидше працюватиме алгоритм);
- необхідності врахування специфіки тер-мінології певної предметної галузі під час підбору варіантів виправлення спотвореного слова;
- обов'язковості формування результату виправлення (коли коректор повинен будь-що сформувати варіанти виправлення);
- критеріїв, за якими оцінюється ефек-тивність роботи коректора;
- значень квазістатичних параметрів ДІС, які відрізняються від тих, що зберігаються в БД агента-коректора.

Крім запитів на обробку спотворених слів, до орфоко-ректора можуть надходити службові повідомлення відносно:

- зміни внутрішнього стану агента-корек-тора;
- надання інформації про стан агента-ко-ректора відправнику даного повідомлення.

У випадку, коли до агента-коректора мо-жуть звертатися кілька модулів ДІС, особливої ваги набуває інформація про його внутрішній стан. Основною характеристикою внутрішньо-го стану є доступність коректора – можливість ініціювання його роботи іншим модулем ДІС. Індикатором доступності може бути як від-сутність оперативних даних, що вже оброблю-ються коректором, так і завантаженість буферів окремих його блоків.

На основі розглянутих вище особливостей будови та функціонування ДІС можна запро-понувати основні способи структурно-алгорит-мічної організації агента-орфоко-ректора.

Організація орфоко-ректора у формі реактив-ного агента

Узагальнена структура програмного агента містить такі компоненти, як [6]: інтерфейс із зовнішнім середовищем, база даних, плану-вальник дій, виконавчий блок. Запропонуємо варіант побудови кожного з цих компонентів агента у випадку, коли останній виконує роль орфоко-ректора (рис. 2).

Інтерфейс із зовнішнім середовищем скла-дається з набору сенсорів та ефекторів, буферів та модуля синтаксичного аналізу/синтезу по-відомлень.

Сенсори та *ефектори* є функціональними блоками, які відповідають за обмін повідом-леннями агента із зовнішнім для нього середо-вищем. При цьому слід зазначити, що в разі віддаленого розміщення компонентів середови-ща вся взаємодія з ними та інші задачі, пов'я-зані з передачею даних, не входять до задач інтерфейсу агента і мають забезпечуватися за-собами комунікації нижчого рівня.

Наявність *буферів* в інтерфейсі, що розгля-дається, дає змогу накопичувати та впорядко-вувати за певною ознакою повідомлення, які надходять із зовнішнього середовища до аген-та, а також повідомлення, сформовані останнім для відправлення іншим модулям ДІС. Це забезпечує надійну роботу агента у випадку, коли до нього звертаються одночасно кілька модулів інформаційної системи.

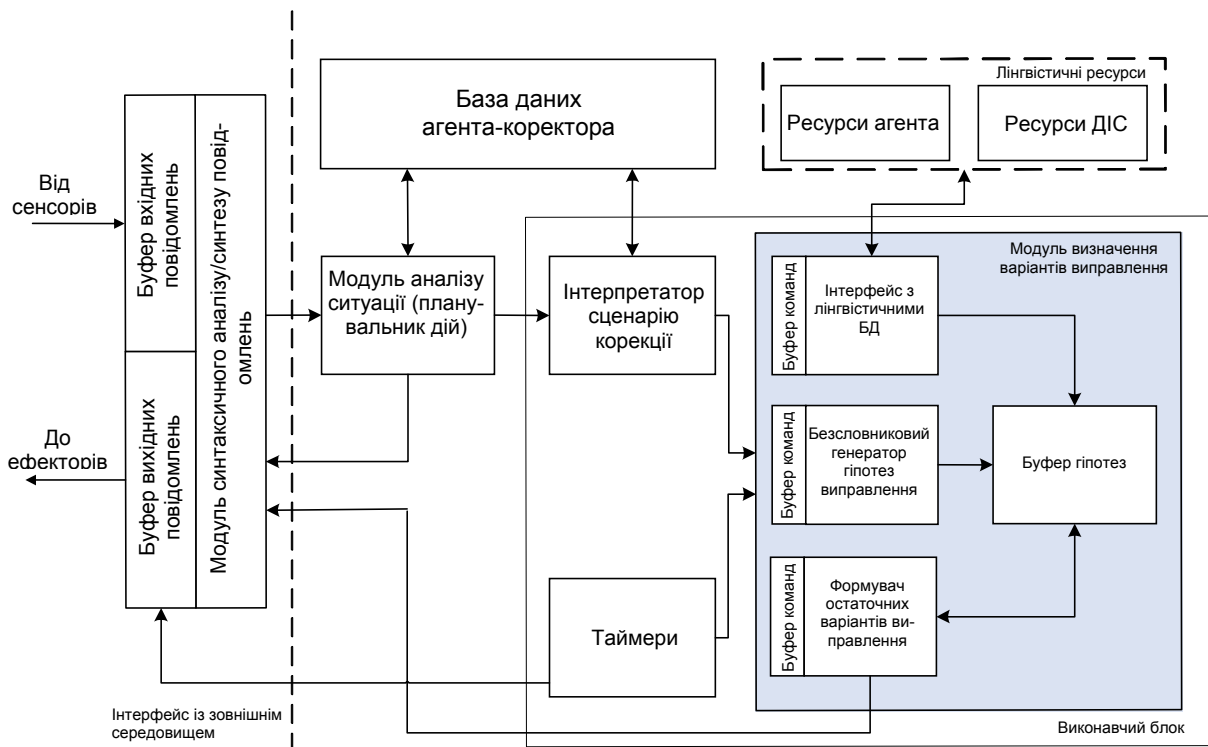


Рис. 2. Структурна схема реактивного агента-коректора

Модуль синтаксичного аналізу/синтезу повідомлень виконує первинний розбір вхідних повідомлень та генерує вихідні повідомлення агента згідно з форматом обміну даними, який прийнятий між компонентами ДІС.

База даних містить записи про поточний стан зовнішнього середовища (значення квазі-статичних параметрів ДІС), про внутрішній стан агента (зокрема, про його доступність), а також зберігає бібліотеки ситуацій (комбінацій значень параметрів, що характеризують стан ДІС та агента), на які здатний реагувати агент-коректор, загальні сценарії поведінки та бібліотеки окремих дій агента. Під сценарієм мається на увазі опис алгоритму роботи агента на високому рівні. Модифікація вмісту БД може виконуватися відповідно до оперативних повідомлень про зміни стану зовнішнього середовища, а також згідно з директивними повідомленнями користувача.

Планувальник дій визначає поведінку агента в реальному часі. Оскільки вище ми охарактеризували орфокоректор як реактивний компонент ДІС, то у нього, звісно, відсутні функції, притаманні агентам, які утворюють та узгоджують цілі (наприклад, інтелектуальний аналіз поточної ситуації, довгострокове планування дій, прогнозування зміни стану зовнішнього

середовища та свого власного стану тощо), що сприяє підвищенню його швидкодії.

Для визначення алгоритму роботи агента-коректора достатньо одного модуля – *модуля аналізу ситуації*, який на основі оперативних даних і наявної в БД інформації проводить вибір сценарію подальших дій агента-коректора. Саме цей модуль повинен враховувати всі розглянуті вище чинники, які впливають на вибір алгоритму роботи коректора. У разі, коли агент не може взяти до обробки вхідне повідомлення, відповідне сповіщення про це має одразу передатися до вихідного буфера. Сценарій подальших дій агента при цьому не вибирається.

Виконавчий блок відповідає за реалізацію вибраного сценарію поведінки агента-коректора. До складу виконавчого блока пропонується ввести *інтерпретатор сценарію*, який здійснюватиме розбір сценаріїв по конкретних командах, та *модуль визначення варіантів виправлення*, який забезпечуватиме виконання цих команд.

У структурній організації агента-коректора вважаємо за доцільне наділити буфером команд кожний модуль, функції якого пов'язані з роботою з варіантами виправлення спотвореного слова. Така особливість будови виконавчого блока забезпечить надійне функціонуван-

ня агента-коректора у випадку, коли інтенсивність потоку запитів до нього є високою. Крім того, при розміщенні модулів виконавчого блока, здатних працювати незалежно один від одного, на окремих комп'ютерах (або в межах багатопроекторної системи), наявність у кожного з них власного буфера команд сприятиме досягненню вищої швидкодії роботи коректора в цілому.

Отже, команди, які містять інструкції щодо виправлення помилок у спотвореному слові, заносяться до відповідних буферів елементів модуля визначення варіантів виправлення з певними службовими позначками, за якими можна однозначно визначити команду та сценарій, до якого вона належить. Решта команд, пов'язаних із внесенням змін до вмісту БД, мають одразу передаватися до СУБД у вигляді запитів.

Агент-коректор повинен також реалізувати механізм ініціювання перевірки буферів своїх складових. Оскільки до структури орфокоректора входить кілька буферів, то здійснення послідовного аналізу їх вмісту уявляється неефективним. Тому необхідним елементом будови агента-коректора є набір таймерів, кожний з яких відповідає за ініціацію перевірки відповідного буфера. Звідси, алгоритм роботи орфокоректора, організованого на основі агентного підходу, зводиться до алгоритму обробки подій таймерів для кожного буфера.

Для забезпечення стабільної роботи виконавчого блока та коректора в цілому потрібно передбачити можливість впливу на них іззовні (користувачем) з метою усунення будь-яких порушень нормального функціонування. Через те що активним компонентом агента-коректора є набір таймерів, потрібно щонайменше реалізувати можливість перезапуску цих таймерів за директивним повідомленням від користувача.

Виходячи із задачі формування множини варіантів виправлення, до складу модуля визначення варіантів виправлення пропонується ввести безсловниковий генератор гіпотез виправлення, інтерфейс з лінгвістичними БД, буфер гіпотез та формувач остаточних варіантів виправлення.

Виділення перших двох елементів зумовлене існуючими варіантами організації процесу автоматизованої корекції спотворених слів: безсловниковим та словниковим. Залежно від особливостей будови та функціонального профілю ДІС агент-коректор може вибрати алгоритм роботи, який передбачає використання або од-

ного з цих елементів, або обох одразу. Крім того, оскільки безсловникові та словникові способи визначення варіантів виправлення спотвореного слова є взаємно незалежними, при розміщенні коректора в межах багатопроекторної системи можливою є організація паралельної роботи безсловникового генератора гіпотез та компонента, який відповідає за підбір варіантів корекції із словникових ресурсів [4].

Для того щоб агент-коректор мав змогу підтримувати взаємодію з різними форматами словникових ресурсів, пропонується в межах алгоритмів роботи відповідного інтерфейсу модуля визначення варіантів виправлення оперувати абстрактними запитами (наприклад, "вибрати множину слів" за критерієм довжини, першої літери, певною граматичною характеристикою тощо), які будуть конкретизуватися за допомогою використання окремих бібліотек, що містять правила синтаксису мови керування даними та структури таблиць БД певного словника. Це дасть змогу налаштувати орфокоректор на роботу з новим словниковим ресурсом за допомогою створення та підключення до нього потрібної бібліотеки.

Безсловниковий генератор гіпотез виправлення повинен містити правила побудови слів мови, якою складені текстові дані, оброблювані ДІС, та мати можливість змінювати набір цих правил у разі потреби.

Отриманий будь-яким чином набір слів, які є кандидатами на виправлення спотвореного слова, потрапляє до буфера гіпотез і знаходиться там доти, доки не буде звідти вилучений і оброблений формувачем остаточних варіантів виправлення. Необхідність внесення до структури орфокоректора буфера гіпотез зумовлена тим, що результати обробки команд, згенерованих інтерпретатором сценарію, можуть надходити від відповідних складових блока орфокорекції не в тій послідовності, в якій їх було відправлено на виконання. Це пояснюється різною швидкістю елементів блока орфокорекції.

Коли в буфері гіпотез є всі необхідні дані, формувач остаточних варіантів виправлення здійснює фільтрацію підібраних слів за додатковими критеріями, що сприяє забезпеченню точності корекції спотвореного слова. Від інтерпретатора сценарію формувач отримує дані про те, який саме модуль має надати набір гіпотез у кожному конкретному випадку і які додаткові фільтри потрібно до цих даних застосувати. Результуюча множина варіантів ви-

правлення передається через синтаксичний аналізатор до вихідного буфера повідомлень.

До набору *програмних словникових ресурсів*, які використовує агент-коректор, непотрібно залучати всі словники, які можуть знадобитися при роботі з різними типами ДІС, адже певна їх частина завжди буде не використана. Тому пропонується ввести до структури орфокоректора лише необхідний мінімум словникових ресурсів, які могли б підтримувати його працездатність, а складні та масштабні словники підключати як компонент ДІС, до якого матимуть доступ й інші модулі системи.

Для того щоб агент-коректор міг обробляти текстові дані різного характеру та реалізувати алгоритми виправлення кількох ступенів складності, до набору словників, якими він користується, повинні входити ресурси різного типу: як прості лексичні словники, що містять просто список слів, так і семантичні ресурси та словники спеціалізованої термінології.

Зазначимо, що варіант організації орфокоректора у вигляді єдиного реактивного агента можна використовувати і у випадку, коли він є автономною програмою. Наявність усіх необхідних ресурсів, а також складових для їх аналізу дозволить коректору працювати в автономному режимі. Внесення змін до БД агента та виконання ініціювання його роботи в такому випадку здійснюватиметься надсиланням користувачем відповідних інструкцій до сенсорів агента.

Організація орфокоректора у формі сукупності агентів

Окремого розгляду потребує випадок, коли коректор вбудовується в ДІС, словникові ресурси якої розміщені на різних комп'ютерах, або коли він реалізує складний ресурсоємний алгоритм пошуку варіантів виправлення спотвореного слова. Якщо при цьому орфокоректор організувати як реактивний агент, ефективність його роботи може знизитися, оскільки швидкість виправлення помилок спаде через втрату часу на передачу даних каналом зв'язку або через нестачу обчислювальної потужності одного комп'ютера. Тому пропонується компоненти модуля визначення варіантів виправлення, які відповідають за формування гіпотез, побудувати у формі допоміжних агентів, кожний з яких виконуватиме певний етап обробки текстових даних та працюватиме з окремим словниковим ресурсом (див. рис. 1, в).

У такому випадку в структурі агента, до якого звертаються модулі ДІС, мають залишитися складові, пов'язані безпосередньо з виправленням помилок, такі, як буфер гіпотез та формувач остаточних варіантів виправлення. Такий агент буде відігравати роль супервізора по відношенню до згаданої групи допоміжних агентів: на основі даних про стан цих агентів він буде вибирати загальний алгоритм обробки запитів на корекцію, а також здійснюватиме формування повідомлень будь-якого характеру до цих програмних елементів розподіленого орфокоректора. Але при цьому взаємодія агента-супервізора з іншими модулями ДІС має залишитися на реактивному рівні.

Координацію роботи агентів, які складають орфокоректор, можна здійснювати за принципом, який відрізняється від механізму спільної роботи частин ДІС, якщо до окремих допоміжних агентів не буде звертатися будь-який інший модуль ДІС. В іншому разі агентна система орфокорекції повинна підтримувати прийнятий у ДІС формат взаємодії компонентів. Структуру агентів, які відповідають за визначення гіпотез виправлення, пропонується зробити подібною до структури реактивного агента-коректора (див. рис. 2). Зазначимо також, що допоміжні агенти в разі необхідності можуть дублюватися в складі орфокоректора.

Висновки

У статті розглянуто сучасний стан проблеми розробки програмних орфокоректорів як складової документоорієнтованих інформаційних систем. Запропоновано в основу структурно-алгоритмічної організації такої утиліти покласти агентний підхід як такий, що широко використовується для реалізації відкритих систем. Вибрано абстрактну архітектуру агента-коректора та визначено, який вплив на його будову та механізми роботи має зовнішнє середовище (ДІС).

Розроблено розподілений та локальний способи структурно-алгоритмічної організації орфокоректора, якими передбачено можливість як спільної роботи з різними типами документоорієнтованих систем, так і інтерактивної роботи лише з користувачем. Розміщення окремих складових агента-коректора на різних процесорних блоках (на різних вузлах мережі) дає можливість паралельно виконувати незалежні етапи процесу визначення варіантів виправ-

лення спотвореного слова, що підвищує загальну швидкодію орфокоректора.

Більш детального вивчення потребує питання сумісного використання лінгвістичних ресурсів документоорієнтованою системою та агентом-коректором. Зокрема, необхідно створити механізм налаштування алгоритмічної

складової орфокоректора на роботу з різними словниковими базами даних. Також перспективним напрямом подальшого дослідження проблеми побудови агентів-орфокоректорів є розробка протоколів взаємодії реактивних програмних агентів, які входять до складу коректора з розподіленою структурою.

Т.Н. Заболотня, А.Ю. Михайлюк, В.П. Тарасенко
СТРУКТУРНО-АЛГОРИТМИЧЕСКАЯ ОРГАНИЗАЦИЯ ПРОГРАММНОГО АГЕНТА-ОРФОКОРРЕКТОРА

Обоснована целесообразность разработки орфо-корректора на основе технологии программных агентов. Определено влияние на строение и механизмы работы агента-корректора внешней по отношению к нему среды – документоориентированной информационной системы. Предложены способы структурно-алгоритмической организации программного корректора для обеспечения его эффективного функционирования в составе документоориентированных компьютерных систем разных типов.

T. Zabolotnia, A. Mykhailyuk, V. Tarassenko

THE STRUCTURAL ALGORITHMIC ORGANIZATION OF THE SOFTWARE CORRECTOR AGENT OF SPELLING ERRORS

The present paper validates the appropriateness of the spelling errors corrector development, based on the software agents' technology. In particular, we determine the influence of the proofreader external environment (document-centric information system) on its structure and working mechanisms. Finally, we apply the means of the structural-algorithmic organization of the software proofreader to ensure its effective functioning as a part of different types of document-centric software systems.

1. *Бугайченко Д.Ю., Соловьев И.П.* Абстрактная архитектура интеллектуального агента и методы ее реализации // Системное программирование. – 2005. – № 1. – С. 36–67.
2. *Тарасенко В.П., Михайлюк А.Ю., Заболотня Т.М.* Специализовані інтелектуальні агенти як засіб інтеграції гетерогенного програмного забезпечення // Інформаційні технології та комп'ютерна інженерія. – 2006. – № 3(7). – С. 96–101.
3. *Филатов В.А.* Мультиагентные технологии интеграции гетерогенных информационных систем и распределенных баз данных: Дис. ... докт. техн. наук: 05.13.06. – Харьков: Харьков. нац. ун-т радиоэлектроники, 2004. – 341 с.
4. *Михайлюк А.Ю., Заболотня Т.М.* Комбінований метод виправлення орфографічних помилок у текстових даних // Вісник Хмельницького нац. ун-ту. – 2007. – 2, № 2. – С. 21–26.
5. *Леонтьева Н.Н.* Автоматическое понимание текстов: системы, модели, ресурсы. – М.: Изд. центр "Академия", 2006. – 304 с.
6. *Кльшинский Э.С.* Агентные системы: классификация и применение // САПР и графика. – 1999. – № 8. – С. 90–96.

Рекомендована Радою
факультету прикладної математики
НТУУ "КПІ"

Надійшла до редакції
20 грудня 2007 року