

СТРУКТУРНО-АЛГОРИТМІЧНА ОРГАНІЗАЦІЯ ПРОГРАМНОГО АГЕНТА-ОРФОКОРЕКТОРА

Вступ

На сьогоднішній день до числа найбільш затребуваних класів програмного забезпечення відносяться 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, котрі входять або мають увійти до складу відкритих ІС, повинні бути побудованими як програмні агенти [1-3].

У даній статті розглядається проблема побудови одного з найпоширеніших інструментів для обробки текстових даних – орфокоректора. На сьогоднішній день він найчастіше розробляється або як жорстко вбудований до ДІС модуль, який здійснює виправлення помилок, або як автономна програма орфокорекції, до котрої при необхідності звертається сам користувач. Обидва наведені способи організації коректорів, по-перше, 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

З огляду на вищезазначене для спрощення інкапсуляції та ефективного функціонування орфокоректора у 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 у 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, і 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 (далі – Система або ДІС), до складу якої він входить, та програмно-апаратне забезпечення цієї системи. Стан даного середовища пропонується визначати за набором значень квазістатичних параметрів ДІС та за оперативними даними, які надходять до агента-коректора у вигляді повідомлень від інших компонентів Системи або від користувача.

Квазістатичними є такі параметри Системи, значення яких під час її роботи змінюються дуже рідко (якщо змінюються взагалі), до числа котрих, перш за все, належать:

- а) структурна організація (будова) Системи - локальне чи віддалене розміщення її компонентів; кількість модулів, які можуть ініціювати роботу агента-коректора; наявність серед програмних ресурсів, котрі використовує Система, таких, які можна залучати до процесу визначення варіантів виправлення помилкового слова тощо;
- б) функціональні характеристики Системи:
 - ступінь складності (інтелектуальності) алгоритмів аналізу текстових даних Системою – визначає загальну необхідність проведення синтактико-семантичного узгодження варіантів виправлення спотвореного слова із його контекстним оточенням чи будь-якими іншими даними;
 - характер текстових даних, з якими працює Система, – формалізований текст, текст, у якому переважають числові дані, структурований текст, художній текст, спеціалізована термінологія тощо;
 - пріоритетність критеріїв визначення ефективності роботи Системи – надання переваги максимальному ступеню автоматизації процесів обробки даних, точності отримуваних результатів, швидкості виконання алгоритмів тощо;
- в) програмно-апаратне забезпечення роботи Системи (і агента-коректора зокрема) - обчислювальна потужність апаратних ресурсів; пропускна здатність каналу передачі даних у випадку віддаленого розташування модулів Системи; параметри операційної системи, під керуванням якої працює ДІС; можливості СУБД, які використовуються для збереження даних тощо.

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 2 види: службові повідомлення та

запити на виправлення спотвореного слова із набором вказівок щодо особливостей проведення процесу корекції.

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

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

Вибір алгоритму роботи агента-коректора виключно на основі аналізу стану зовнішнього середовища не завжди гарантує його ефективного функціонування у межах багатокomпонентної відкритої ДІС. Для забезпечення більш гнучкої та надійної роботи орфококоректора пропонується будувати його як агент, 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 орфококоректора, які він змінює під впливом стану зовнішнього середовища. Пропонується до множини I включити дані про поточний стан складових агента-коректора, а також дані про можливість ініціювання роботи орфококоректора 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 орфококоректор є сервісним доповненням Системи, 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 є вкрай небажаним. Таким чином, по відношенню до Системи агент-коректор має бути реактивним.

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

Структурна організація документоорієнтованої ІС. У випадку *локального* розміщення компонентів ДІС орфококоректор доцільно реалізувати у формі одного *реактивного* агента (див. рис 1а) [3]. При цьому компактність Системи 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 Системи відображений перерваною лінією).

Побудова коректора у вигляді сукупності реактивних агентів Agt_i (наприклад, для забезпечення паралельного виконання певних незалежних етапів алгоритму виправлення помилок [4]) має сенс тільки для Системи, що функціонує на багатопроцесорному комп'ютері.

Форма реалізації орфококоректора в умовах *віддаленого* розміщення складових ДІС залежить від взаємного розташування модулів (M_i) та ресурсів (Res_i), з котрими він взаємодіє в процесі визначення варіантів виправлення спотвореного слова. Чим більше компонентів, необхідних для роботи коректора, знаходиться на одному з них вузлі комп'ютерної мережі, тим швидше будуть отримані результати корекції і тим менше даних буде передаватися при цьому каналом зв'язку (див. рис 1б). Випадок, коли всі потрібні компоненти належать одному вузлу, можна розглядати як локальне розміщення Системи та коректора, і, відповідно, реалізувати останній 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), допускається організація орфококоректора у формі *багатоагентної* системи, кожна складова якої буде відповідати за організацію роботи з окремим ресурсом або реалізовувати певний етап процесу корекції помилок у спотвореному слові (див. рис 1в). При цьому агент Agt_{main} , який безпосередньо взаємодіє з ДІС, з одного боку має контролювати роботу усіх інших агентів Agt_i , але, водночас, залишатися реактивним по відношенню до компонентів ДІС.

Також потрібно звернути увагу на те, що звертання коректора до ресурсів Системи може спричинити зниження її загальної надійності та швидкодії.

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

Порівняно із локальним варіантом розміщення Системи при віддаленому розташуванні її компонентів зростає імовірність 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 Системи може здійснювати моніторинг зайнятості агента-коректора, або орфококоректор має самостійно контролювати спроби одночасного звернення до нього з боку декількох модулів Системи.

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

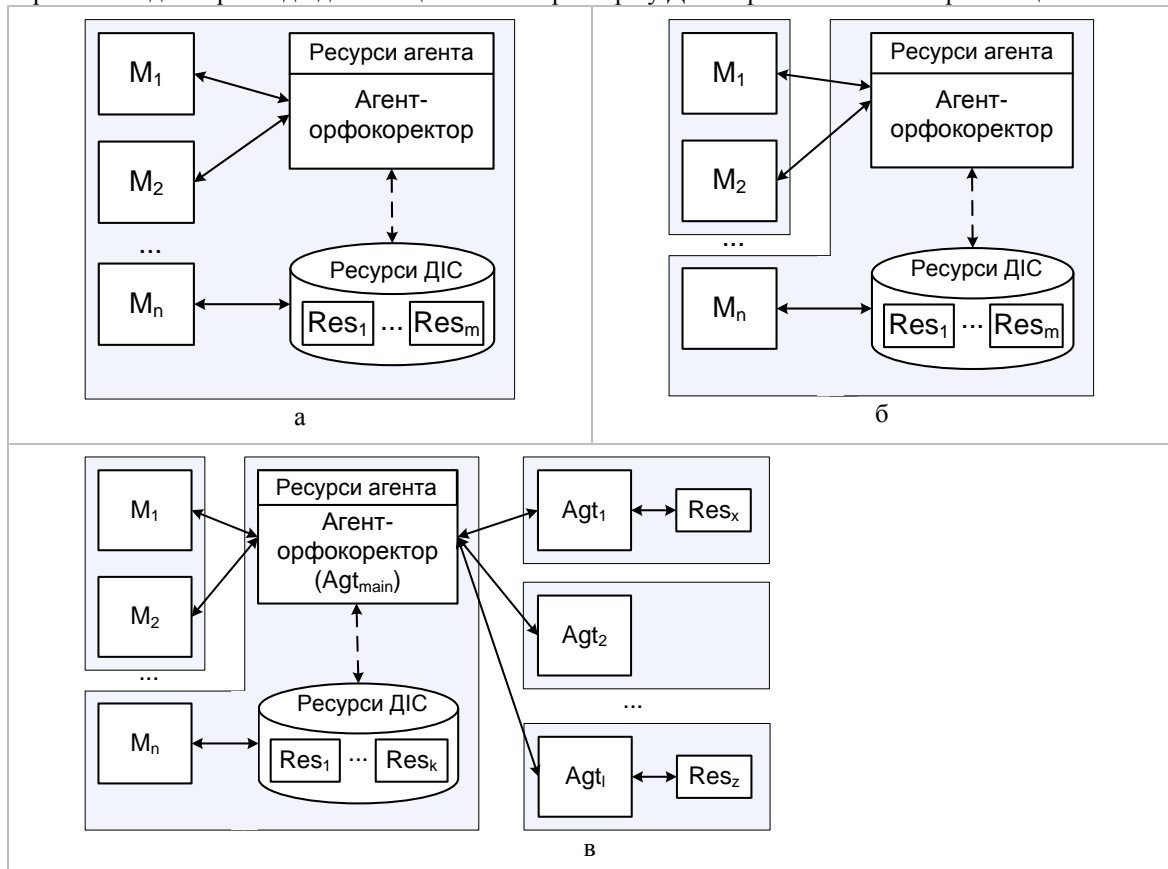


Рис. 1. Варіанти реалізації агента-коректора у залежності від розташування компонентів ДІС

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

Ступінь складності алгоритмів обробки текстових даних. На сьогоднішній день існуючі документоорієнтовані системи та інше ПЗ автоматизованої обробки текстів [5] у залежності від свого призначення реалізують різні за складністю алгоритми аналізу текстових даних. Можна стверджувати, що у переважній більшості випадків ступінь складності алгоритмів обробки спотвореного слова відповідним орфококоректором залежить від рівня аналізу текстів, який має місце у Системі в цілому. Так для програми, яка оперує даними на рівні морфологічного аналізу тексту (наприклад, для простого редактора тексту), не потрібно будувати коректор, що буде узгоджувати варіанти виправлення зі змістом контекстного оточення слова, у 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 орфококоректора 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 орфококоректором.

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 у документах, а також предметноорієнтованих словникових ресурсів, які дозволили б адекватно ідентифікувати та виправити спотворене слово. Корекція формалізованого тексту (програмного коду або документу, 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 даних

формалізованій мові. Для знаходження вірного варіанту написання спотвореного слова з художнього тексту необхідне проведення детального синтактико-семантичного аналізу контексту даного слова із використанням словників, які містять інформацію про слова, що можуть мати переносне значення, усталені мовні звороти та інші прийоми художньої виразності. Якщо ж текст, у якому здійснюється орфокорекція, насичений числовими даними, застосування алгоритмів та ресурсів, які використовуються для виправлення помилок у словах, не є доцільним. Замість цього потрібно використовувати способи знаходження спотворень у послідовностях цифр, які мають місце у галузі кодування.

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

Програмно-апаратне забезпечення роботи ДІС. Від характеристик апаратних ресурсів та програмного забезпечення, на базі яких функціонує ДІС, залежить припустима складність алгоритмів обробки текстових даних (у т.ч. орфокорекції), обсяг та структурна організація даних у використовуваних ресурсах, принципова можливість віддаленого розташування 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, мають 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, вище було розглянуто вплив значень квазістатичних параметрів ДІС на структурно-алгоритмічну організацію агента-коректора як складової цієї системи. Оскільки ці значення змінюються дуже рідко, 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 у ДІС та оновлювати відповідно до змін у Системі.

Оперативні повідомлення від ДІС. Для того, щоб ініціювати роботу орфокоректора, передати йому необхідні текстові дані, а також додаткові вказівки (інструкції) щодо процесу виправлення помилок, відповідний модуль ДІС чи користувач мають сформулювати певне оперативне повідомлення і надіслати його до орфокоректора. 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.

Переважно більшість оперативних повідомлень складають безпосередні запити на виправлення спотвореного слова із набором допоміжних інструкцій щодо проведення процесу корекції. Кожен запит умовно поділяється на 2 частини: основні дані та додаткові інструкції. До основних даних належить спотворене слово та його контекст (у разі, коли потрібно здійснювати синтактико-семантичний аналіз контекстного оточення).

Інструкції можуть містити дані щодо:

- максимальної кількості помилок, яку має виправляти коректор (чим менше помилок, тим швидше буде працювати алгоритм);
- необхідності врахування специфіки термінології певної предметної галузі під час підбору варіантів виправлення спотвореного слова;
- обов'язковості формування результату виправлення (коли коректор повинен будь-що сформулювати варіанти виправлення);
- критеріїв, за якими оцінюється ефективність роботи коректора;
- значень квазістатичних параметрів ДІС, які відрізняються від тих, що зберігаються у БД агента-коректора.

Крім запитів на обробку спотворених слів до орфокоректора можуть надходити службові повідомлення:

- щодо зміни внутрішнього стану агента-коректора;
- щодо надання інформації про стан агента-коректора відправнику даного повідомлення.

У випадку, коли до агента-коректора можуть звертатися декілька модулів ДІС, особливої ваги набуває інформація про його внутрішній стан. Основною характеристикою внутрішнього стану є доступність коректора

- можливість ініціювання його роботи іншим модулем ДІС. Індикатором доступності може бути як відсутність оперативних даних, що вже оброблюються коректором, так і завантаженість буферів окремих його блоків.

На основі розглянутих вище особливостей будови та функціонування ДІС можна запропонувати основні способи структурно-алгоритмічної організації агента-орфокоректора.

Організація орфокоректора у формі реактивного агента

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

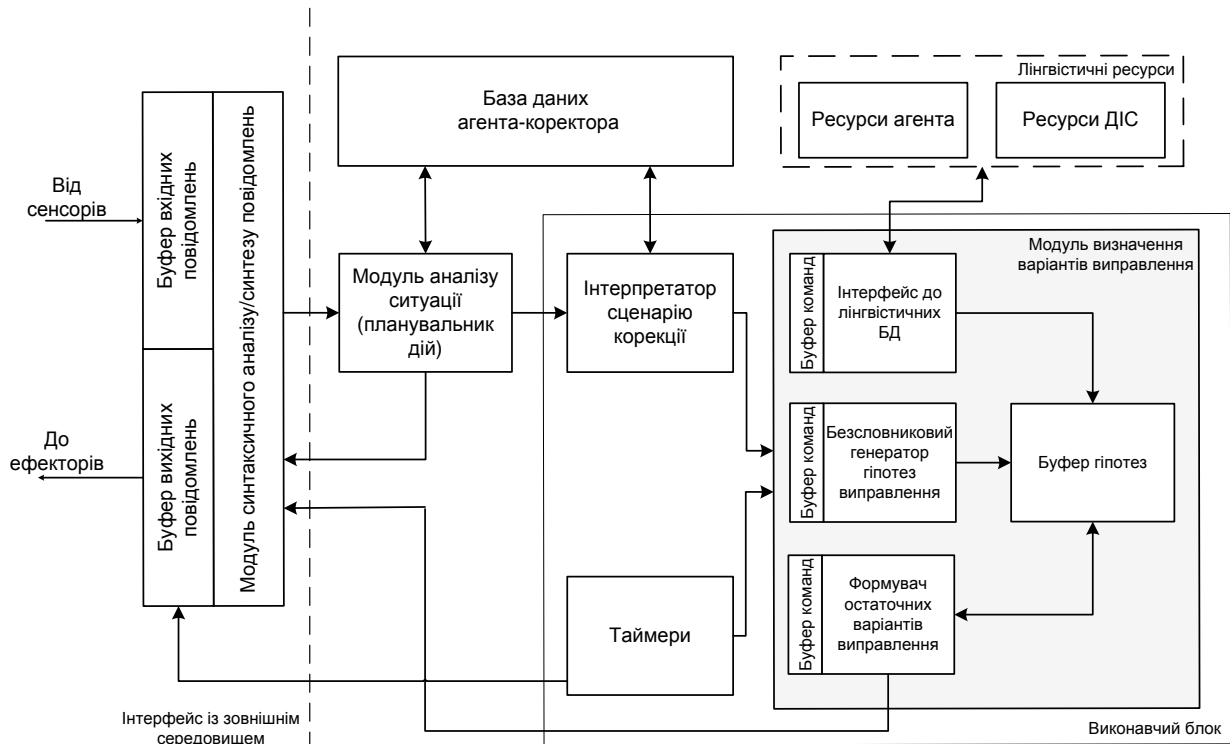


Рис. 2. Структурна схема реактивного агента-коректора

Інтерфейс із зовнішнім середовищем складається з набору сенсорів та ефекторів, буферів та модуля синтаксичного аналізу/синтезу повідомлень.

Сенсори та ефектори є функціональними блоками, які відповідають за обмін повідомленнями агента із зовнішнім для нього середовищем. При цьому слід зазначити, що у разі віддаленого розташування компонентів середовища вся взаємодія із ними та інші задачі, пов'язані із передачею даних, не входять до завдань інтерфейсу агента і мають забезпечуватися засобами комунікації більш низького рівня.

Наявність **буферів** у інтерфейсі, що розглядається, дозволяє накопичувати та упорядковувати за певною ознакою повідомлення, які надходять із зовнішнього середовища до агента, а також повідомлення, сформовані останнім для відправлення іншим модулям ДІС. Це забезпечує надійну роботу агента у випадку, коли до нього звертаються одночасно декілька модулів інформаційної системи.

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 квазістатичних параметрів ДІС), про внутрішній стан агента (зокрема про його доступність), а також зберігає бібліотеки ситуацій (комбінацій значень параметрів, котрі 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, у нього 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 прийняти до обробки вхідне повідомлення, відповідне сповіщення про це має одразу передатися до вихідного буфера. Сценарій подальших дій агента при цьому не обирається.

Виконавчий блок відповідає за реалізацію обраного сценарію поведінки агента-коректора. До складу виконавчого блоку пропонується ввести *інтерпретатор сценарію*, який буде здійснювати розбір сценаріїв по конкретних командах, та *модуль визначення варіантів виправлення*, котрий буде забезпечувати виконання цих команд.

У структурній організації агента-коректора вважаємо за доцільне наділити буфером команд кожний модуль, функції якого пов'язані із роботою з варіантами виправлення спотвореного слова. Така особливість будови виконавчого блоку забезпечить надійне функціонування агента-коректора у випадку, коли інтенсивність потоку запитів до нього є високою. Крім того, при розміщенні модулів виконавчого блоку, здатних працювати незалежно один від одного, на окремих комп'ютерах (або у межах багатопроцесорної 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 із 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, повинні 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, 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 блоку та коректора в цілому потрібно передбачити можливість впливу на них ззовні (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 включити *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. У залежності від особливостей будови та функціонального профілю ДІС агент-коректор може обрати алгоритм роботи, який передбачає використання або одного з цих елементів, або обох одразу. Крім того, оскільки безсловникові та словникові способи визначення варіантів виправлення спотвореного слова є взаємно незалежними, при розташуванні коректора у межах багатопроцесорної 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 у 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 управління даними та структури таблиць БД певного словника. Це дозволить налаштувати орфокоректор на роботу із новим словниковим ресурсом шляхом 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 шляхом набір слів, котрі 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 включення до структури орфокоректора буфера гіпотез спричинена тим, що результати обробки команд, згенерованих інтерпретатором сценарію, можуть надходити від відповідних складових блоку орфокорекції не в тій послідовності, в якій їх було відправлено на виконання. Це пояснюється різною швидкістю елементів блоку орфокорекції.

Коли у 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. Результируюча множина варіантів виправлення передається через синтаксичний аналізатор до вихідного буфера повідомлень.

До набору **програмних словникових ресурсів**, які використовує агент-коректор, не потрібно включати всі словники, які можуть знадобитися при роботі з різними типами ДІС, адже певна їх частина завжди буде невикористана. 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 реалізовувати алгоритми виправлення декількох ступенів складності, до набору словників, якими він користується, повинні входити ресурси різного типу: як прості лексичні словники, котрі 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 у автономному режимі. Внесення змін до БД агента та виконання ініціювання його роботи у такому випадку здійснюватиметься шляхом надсилання користувачем відповідних інструкцій до сенсорів агента.

Організація орфокоректора у формі сукупності агентів

Окремого розгляду потребує випадок, коли коректор вбудовується до ДІС, словникові ресурси якої розташовані 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 впаде через втрату часу на передачу даних каналом зв'язку або через нестачу обчислювальної потужності одного комп'ютера. Тому пропонується компоненти модуля визначення варіантів виправлення, котрі 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. 1в).

У такому випадку у 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 обирати 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 повинна залишитися на реактивному рівні.

Координацію роботи агентів, котрі 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 Системи. У іншому разі агентна система орфокорекції повинна підтримувати прийнятій у ДІС формат взаємодії компонентів. Структуру агентів, які відповідають за визначення гіпотез виправлення, пропонується зробити подібною до структури реактивного агента-коректора (див.рис.2). Також зазначимо, що допоміжні агенти в разі необхідності можуть дублюватися у складі орфокоректора.

Висновки

Розглянуто сучасний стан проблеми розробки програмних орфокоректорів як складової документо-орієнтованих інформаційних систем. Запропоновано в основу структурно-алгоритмічної організації такої утиліти покласти агентний підхід як такий, що широко використовується для реалізації відкритих систем. Обрано 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 із різними типами документо-орієнтованих систем, так і інтерактивної роботи лише з користувачем. Розташування окремих складових агента-коректора на різних процесорних блоках (на різних вузлах мережі) дозволяє паралельно виконувати незалежні етапи процесу визначення варіантів виправлення помилкового 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.

Список літератури

1. *Бугайченко Д.Ю., Соловьев И.П.* Абстрактная архитектура интеллектуального агента и методы её реализации // Системное программирование. – СПб.: СПбГУ, 2005, №1. – С. 36–67.
2. *Тарасенко В.П., Михайлюк А.Ю., Заболотня Т.М.* Спеціалізовані 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... д-ра техн. наук: 05.13.06 / Х.: Харьковский национальный ун-т радиоэлектроники, 2004. – 341 с.
4. *Михайлюк А.Ю., Заболотня Т.М.* Комбінований метод виправлення орфографічних помилок у текстових даних // Вісник Хмельницького національного університету. – 2007. – № 2. – Т.2. – С. 21–26.
5. *Леонтьева Н.Н.* Автоматическое понимание текстов: системы, модели, ресурсы. – М.: Издательский центр "Академия", 2006. – 304 с.
6. *Кльшинский Э.С.* Агентные системы: классификация и применение // САПР и графика. – 1999. – № 8. – С. 90–96.