УДК 904.622

Марущак В.І., ***Філатова Т.В.,*** *Яковенко* ***О.В,***
Одеський національний політехнічний університет,
пр. Шевченка 1, Одеса, Україна, 65044

ПІДХІД ДО ЗБОРУ ДАНИХ ПРО СТРУКТУРУ ДОКУМЕНТООБІГУ

Анотація

Одна з найважливіших підготовчих (раз впровадження системи документообігу - збір даних про структуру документообігу. Запропоновано підхід, заснований на автоматизованому аналізі документів, що використовуються у організації. Зібрані дані уточнюються завдяки опитуванню, формалізованому у підході. Також запропоновано моделі документів і потоків робіт для формалізації процесу.

Resume

One of important preparatory phases of workflow system implementation is collecting of workflow structure data. We propose an approach based on automatized analysis of documents used in the organization. Gathered data must be specified using interrogations formalized by the approach. Also we propose a model of documents and a model of workflow data to formalize the process.

Вступ

Упровадження систем електронного документообігу (СЕД) - складний процес, що потребує великих витрат часу і ресурсів. При цьому є значна ймовірність того, що процес впровадження закінчиться невдало: нерідко стається так, що структура електронного документообігу не відповідає реальним процесам, які виконуються у організації, і введена до експлуатації система залишається незатребуваною.

Отже, збір інформації про електронний документообіг є критичною частиною впровадження СЕД у організації, що має довготривалий вплив на подальше функціонування системи. Існуючі методи збору не враховують усіх типів даних, що використовуються у сучасних СЕД, тому ми пропонуємо удосконалення сучасних підходів.

Існуючі підходи

Збір інформації про електронних документообіг обумовлюється тою моделлю, що вибрана для представлення документообігу. При цьому під документообігом у більшості випадків розуміється структура потоків робіт, які часто представляються у вигляді потоків документів. Існує декілька відомих варіантів представлення таких потоків і пов'язаних з ними методів проектування.

Історично найбільш поширеним є структурний аналіз, відомий як SADT (Structured Analysis and Design Technique), що базується на стандартах IDEF (Integrated Definition). Особливістю SADT є те, що він однаково широко використовувався і програмістами (комп'ютерними аналітиками), які бажали описати програмні системи, і економістами, що моделювали діяльність організацій [1]. Цей факт і став основою популярності SADT при впровадженні СЕД. Зокрема, опис бізнес-процесів за допомогою IDEF3 є гарною основою для проектування структури електронного документообігу, оскільки представляє собою сукупність блоків, що позначають дії (у випадку СЕД над документами), пов'язаних між собою переходами.

Серед вчених, які бажають моделювати динамічні характеристики документообігу, найбільш поширені мережі Петрі та їх розширення. Мережі Петрі можуть описувати не тільки структуру документообігу, а й розраховувати критичне навантаження та подібні характеристики. У організаціях, що впроваджують системи документообігу, такі моделі не знайшли популярності, оскільки не є очевидними та не можуть вдало описувати структуру документообігу [2].

Інший поширений метод, що може бути використана для аналізу документообігу, - об'єктно-орієнтований аналіз, розроблений Граді Бучем [3]. Незважаючи нате, що його основна направленість - програмні системи, цей метод надзвичайно підходить для дослідження документообігу, а діаграми UML-для опису структури документообігу [4]. Основною перевагою цих діаграм є те, що вони дозволяють представляти одну і ту ж систему документообігу з різних проекцій: діаграма діяльності допомагає докладно описати робочий процес, діаграма взаємодій - обмін даними між модулями, діаграма станів - маршрут документа, діаграма класів - структуру організації.

Для збору даних найчастіше пропонується використовувати опит експертів у підрозділах. Крім того, у діловодстві розроблені так звані табелі для агрегації формальних атрибутів документів.

Особливості підходу до збору даних про електронний документообіг

Сучасні системи документообігу характеризуються розширенням своїх функцій та інтеграцією у рамках у рамках однієї системи декількох різнопланових модулів. Крім того, існуючі методи збору потребують кілька ітерацій для підтвердження зібраних даних, оскільки великий обсяг даних обумовлює значну кількість помилок. Також існуючі методи не враховують різну кваліфікацію тих працівників, що будуть проводити збір інформації.

Ми пропонуємо підхід, що дозволяє організувати поступовий процес збору даних для побудови СЕД за допомогою автоматизованої програмної системи, враховуючи різні типи користувачів та забезпечуючи аналітиків частковою узгодженою інформацію на проміжних етапах процесу.

У процесі аналізу повинні бути підготовлені дані, достатні для побудови мережі документообігу, аналізу зв'язків між документами, аналізу темпоральних характеристик обробки документів, автоматичного контролю робіт, аналізу існуючої інфраструктури організації! побудови оптимального впровадження СЕД.

Аналіз даних, потрібних для побудови СЕД

Очевидним базовим набором даних є інформація про структуру потоків документів, або потоків робіт, що виконуються в організації. Ця підсистема є базової для більшості інших підсистем, що забезпечують контроль документообігу і спільну роботу в організації. Крім даних про структуру потоків, повинна бути зібрана інформація про інтенсивність обробки документів, що характеризує їх вагу у рамках системи.

Таким чином, дані, які треба зібрати у результаті, відносяться до одного з видів: список доїсументів, що використовуються у документообігу, маршрути документів у документообігу, залежності між документами і зв'язки по даним, завантаження підрозділів по періодах, терміни виконання робіт.

Очевидно, що одночасний збір інформації по усім видам не тільки є трудомістким, але і важкий організаційно, оскільки володіти необхідними відомостями можуть різні експерти. Крім того, частина цих даних міститься у самих документах і може бути проаналізована без допомоги експертів. Щоб додати процесу збору інформації гнучкості, доцільно розділити дані на групи: тип документу та його автор; інформація по маршрутам і залежностям між документами, що збирається на основі тексту документа; інформація по маршрутам, що збирається на основі опитування експерта; інформація по залежностям між документами, що збирається на основі опитування експерта; інформація по інтенсивності, або темпоральні характеристики (середній об'єм документа, періодичність термін виконання та інше); список необхідних і створених кластерів даних.

Інформація, що відноситься до певної групи, є цілісною і може бути уведена користувачем у рамках єдиного кроку.

Базовою логічною одиницею системи документообігу є документ, тому ми маємо базувати модель даних саме на цьому об'єкті. Запропонуємо модель даних, що дозволяє описати документ.

Модель документа *d O D є* кортежем $d = <td, dsc, au, p, R, V, s, b, E, G, L>$, де td O *TD* — вид документа, *dsc* - зміст (опис), *au* O *U* автор документа, *s* O *U*-зазначений виконавець, *R* H *U* - множина осіб, що надають затвердження, V H *U*- множина осіб, що надають візу, s O *U*- особа, що підписує, *b* O *U*- особа, що внесла проект, *E* M *D* - додатки, G M *D* - підстави документа, *L* M *U* множина користувачів, що мають отримати документ за допомогою розсилання, *U*- множина користувачів документообігу. За допомогою цієї моделі можливо представити інформацію про документ, його атрибути, залежність між документами.

За допомогою даних, що представлені у вигляді зазначеної моделі документу, можна описати структуру документообігу. Класична структура документообігу може бути представлена як кортеж $wf= <g, G'>$, де $g= \{Node, Link\}$ - граф мережі документообігу, *Node* - множина вузлів мережі, що представляють підрозділи організації, *Link* - множина зв'язків між вузлами, що представляють можливі маршрути руху елементів. $G' = \{g'\}$ - це множина маршрутів документів, де gi' = *(Node* i.', *Link* i'*}, Node* i.' H *Node* - множина вузлів, де обробляється документ *i, IinЬ.'HЬinЬ* - множина можливих переходів для документа;'.

Така модель потоків робіт легка для розуміння, але не є строгою і дозволяє втрати даних. Тому уточнена модель даних, що представляють структуру потоків робіт, представимо у вигляді кортежу $wf'= <AT, J, DS, DW>$, де *AT*- множина типів дій над документами, *J* - множина посад (користувачів), *DS* - множина станів документів, *DW*- опис переміщень документів

Збір даних до моделі м/'зводиться до заповнення бази даних, представленої моделлю потоків робіт. Щоб це реалізувати, між двома моделями установлена відповідність, що дозволяє автоматично перетворювати дані між моделями: $J=f_1(U), AT=f_2(p, b, E, G, L), DS =f_3(p, b, E, G, L), DW=f_2(p, b, E, G, L)$. Реалізація цих функцій виконана у вигляді алгоритму.

Окремо формалізуємо темпоральні характеристики, що дозволяють створити динамічну модель документообігу, їх заповнення не є необхідним для впровадження СЕД, але важливе для моделювання та забезпечення таких додаткових функцій, як автоматична перевірка виконання завдань. Темпоральні характеристики документу є кортежем $d_t = < vt, st, pt, rt>$, де *vt* - середній об'єм документу, *st* - кількість документів у одиницю часу, *pt* - періодичність створення документів, *rt* - офіціальний час виконання. У залежності від типу організації

представлення цих атрибутів може змінюватись від вільного опису до числового вигляду (типового).

Посилання на кластери даних також є необов'язковим, але цей набір даних є незвичним: для його аналізу потрібен не експерт з обробки відповідного документу, а спеціаліст з баз даних. Ця характеристика зобов'язана забезпечити зв'язок документів з реляційною базою даних інформаційної системи, базуючись на тому принципі, що документи переносять ті ж дані, що обробляються додатками баз даних. Кластер даних - набір даних з бази даних, що зібраний до купи на основі логічного змісту даних. Це може бути перелік співробітників, дані про відпустки, дані про надбавки і таке інше. Розділення даних інформаційної системи проводить спеціалістом з баз даних, і тому аналіз документів за допомогою кластерів даних також повинен проводитися спеціалістом. Для визначеності треба створити перелік (множину) кластерів даних $CD = \{cd\}$ та для кожного документа визначати, створює він ці дані чи використовує. Таким чином, посилання на кластери є двома множинами d_{CD} - $\{ CD_{in}, CD_{out} \}$, де CD_{in} - множина кластерів даних, що використовуються документом, CO_{mI} - множина кластерів даних, що створюються документом. Очевидно, що об'єднання $CO._n$ для всіх документів має бути підмножиною об'єднання усіх CD_{out}. Така структура дозволяє використовувати різноманітне тестування для маршрутів документів [5].

Контроль даних

Контроль даних у системі збору інформації про документообіг забезпечується за допомогою надмірностей даних. Ці надмірності бувають двох типів: різні джерела даних чи різні типи даних з одного джерела. Різні джерела даних створюються у залежності від предметної області: наприклад, множина кластерів даних може бути сформована як на основі документів, так і на основі бази даних інформаційної системи. Такі порівняння потребують втручання аналітика, тому більш важлива формалізація порівняння різних типів даних, що були отримані з одного джерела.

Прикладом таких даних, що були отримані з надмірністю, є залежності документів. Для кожного документа заповнюються його підстави та наслідки у вигляді розсилки. Усі ці дані готові до поєднання до єдиної системи: представимо їх у вигляді орієнтованого графа gc - $\{CNode, CLink\}$. Множина вузлів орієнтованого графа є об'єднанням двох множин: множини оброблених документів D_a і множини передачі даних ". Коли ми додаємо до множини документів новий елемент d, ми повинні з'єднати його тою кількістю дуг, скільки підстав і наслідків у нього нараховано. Якщо відповідного вузла не знайдено, до множини " додається новий документ. Кожний елемент з " має мати одну вхідну дугу та мінімум одну вихідну, що перевіряється за допомогою алгоритму повного перебору усіх дуг.

Керування послідовністю внесення даних

Оскільки збір інформації про документообіг є складним і тривалим процесом, навіть частковий набір даних повинен бути корисним при аналізі й впровадженні СЭД. Отже, до процесу збору інформації пред'являється вимога інкрементності, тобто послідовного нарощування набору даних таким чином, що на кожному кроці отримані дані є несуперечливими і дозволяють забезпечити документообіг у частині підрозділів так, щоб покриття було найефективнішим.

При цьому треба врахувати той факт, що робота аналітиків у організації проводиться у контакті з відповідальними особами, що мають власні оцінки важливості обміну даними з окремими підрозділами. Як наслідок, необхідно врахувати їх побажання при формуванні послідовності збору.

Для визначення послідовності збору інформації пропонується модель, призначена для визначення оптимальної послідовності збору інформації в підрозділах за критерієм максимального об'єму даних для обміну, забезпеченого СЕД у кожний момент розвитку. Передбачається, що вибір першого підрозділу робиться користувачем на основі досвіду: у більшості випадків для цього підходить загальний відділ або відділ кадрів.

Нехай sij- інтенсивність створення документа dj що формується на основі документу dj. Нехай D- множина усіх документів, а D_a - множина документів, інформація по яким уже зібрана. Нехай R - матриця інцидентності документів, в якій $r_{ij} = 1$, якщо документу формується на основі документа i. Нехай $b_{li} = 1$, якщо документ D_l обробляється у підрозділі i, та 0 у іншому випадку, а $c_m=1$, якщо документ d_m належить до множини D_a, і 0 у іншому випадку.

Тоді для усіх підрозділів ui, що плануються для обстеження, необхідно розрахувати характеристику >>i

$$\lambda_i = k_i \cdot \sum_{m=1}^{N} \sum_{l=1}^{N} \left(\left(s_l r_{ml} + s_m r_{lm} \right) \cdot b_{li} c_m \right),$$

де k_i - коефіцієнт пріоритету підрозділу i, уведений ззовні, N- кількість документів. Точність даного розрахунку може бути збільшена за допомогою врахування характеристики », розрахованої для вже оброблених документів:

Таким чином, запропоновано формальний підхід до визначення послідовності збору даних у великих організаціях.

Висновки

Запропонований підхід знайшов відображення у програмі для збору даних, що є додатком реляційної бази даних. Формалізація даних дозволила організувати такий процес збору даних, що в ньому змогли взяти участь користувачі з різним рівнем кваліфікації, і тимчасова відсутність експерта з окремого підрозділу не зупиняє роботу. Таким чином, підхід став важливою складовою частиною процесу впровадження СЕД.

Список використаних джерел:

/. ЧеремньїхС.В., СеменовИ.О., РучкинВ.С. Структурний анализ систем: IDEF-технологии. —М.: Финансы и статистика, 2003. - 208 с.

2. Cardoso J., Bostrom R. P., A. Sheth Workflow Management Systems and ERP Systems: Differences Commonalities, and Applications //Information Technology and Management. -2004. -№ 5. -P. 319-338.

3. Буч Г. Обьектно-ориентированньій анализ и проектированне с примерами приложений на C++; *Пер. с англ. - М.: «Бином», 1998 - 560 с.*

4. Ларман К. Применение UML и шаблонов проектирования: Пер. с англ. - М.: Издательский дом «Вильямс», 2004. - 624 с.

5. Орлов С. Технологии разработки программного обеспечения - СПб.: Питер, 2002. - 464 с.