

СТРУКТУРИЗАЦІЯ ДАНИХ В СИСТЕМАХ ЕЛЕКТРОННОГО ДОКУМЕНТООБІГУ

Лозович Т.М.¹⁾, Адамів О.П.²⁾

Тернопільський національний економічний університет

^{1)аспірант; ^{2)к.т.н., доцент}}

I. Постановка проблеми

Інформатизація і комп'ютеризація є основою сучасного розвитку суспільства і головним чинником ефективного здійснення управлінських функцій установи та організації.

Основна функція будь-якої установи – прийняття управлінських рішень. При цьому вихідні дані для прийняття рішень, необхідна інформація і самі рішення представляються у вигляді службових документів.

Введення системи електронного документообігу (СЕД) допомагає оптимізувати роботу організації з документами і керувати ними протягом усього їх життєвого циклу.

Отже, сучасна система електронного документообігу має значні переваги по роботі з документами, однак існує ряд недоліків таких систем. В першу чергу це робота лише із структурованими даними. Таким чином, слабоструктуровані або неструктуровані різнотипні дані, що складаються із документів, в системі електронного документообігу не виконують тих завдань, які забезпечують ефективне функціонування організації.

II. Мета роботи

Метою роботи є аналіз поняття документ, а також його роль в системі електронного документообігу, аналіз структурованих та неструктурованих даних та їх місце в СЕД. Визначення проблем неефективного функціонування електронного документообігу, пов'язаних із поганою структуризацією документів у системі.

III. Особливості структуризації в СЕД

Документ – це базова теоретична конструкція, яка відноситься до всього, що може бути збережене або представлено, щоб служити як доказ для певної мети.

Він є засобом закріплення різними способами на відповідному матеріалі інформації про: факти, події, явища об'єктивної дійсності та розумову діяльність людини.

Видатний бельгійський вчений, автор декількох сотень праць про проблеми теорії та практики бібліографії, бібліотекознавства і документації Поль Отле, характеризує документ як будь-яке джерело інформації, передання людської думки, знань незалежно від того, чи втілене воно в матеріально фіксованій формі або є провідником (передавачем) інформації в часі.

В Україні офіційно прийняті три визначення документа, зафіксовані в деяких Держаних стандартів (ДСТУ):

ДСТУ 2392-94: «Документ - записана інформація, яка може розглядатися як одиниця під час здійснення інформаційної діяльності».

ДСТУ 3017-95: «Документ - матеріальний об'єкт з інформацією, закріпленою створеним людиною способом для її передавання в часі та просторі».

ДСТУ 2732:2004: «Документ - інформація, зафіксована на матеріальному носії, основною функцією якого є зберігати та передавати її в часі та просторі».

У статті 1 Закону України «Про інформацію» говориться, що документ - матеріальний носій, що містить інформацію, основними функціями якого є її збереження та передавання у часі та просторі [1-3].

Таким чином, існує декілька визначень поняття «документ». Усі вони мають спільні та відмінні риси. Однак, при аналізі визначень, нами було запропоновано таке, яке найбільше відповідає предметній області даної статті. Документ – це багатфункціональна інформаційна одиниця, зафіксована на матеріальному носії. Інформаційна одиниця, у вузькому розумінні, файл або група файлів, які розглядаються як одне ціле. Однак, дане визначення стосується лише цифрових об'єктів (електронні документи, фото, відео, аудіо файли тощо). У широкому розумінні, інформаційна одиниця – це неподільний елемент даних, або їх набір, який розглядається як одне ціле. Поняття «подільної» та «неподільної» інформації є досить умовним, оскільки не існує чітких критеріїв, які б віднесли об'єкт до тієї чи іншої категорії. Проте, прийнято вважати, що документ – це цілісний

об'єкт, поділ якого є практично неможливий, оскільки при цьому втрачається його інформаційна значимість.

В сучасному інформаційному суспільстві, будь-яка організація стикається з величезною кількістю вхідних та вихідних даних. Документи, з якими вони ведуть роботу, це не тільки друкований чи писаний текст з певними реквізитами, для їх юридичної сили, але й різного роду внутрішня та зовнішня інформація, яка може бути представлена як текстом, так і за допомогою мультимедіа. Організації змушені працювати з цими даними, оскільки це питання стосується їхньої інформаційної безпеки та іміджу.

Отже, сучасне інформаційне суспільство вимагає об'ємної роботи, пов'язаної із великим потоками даних. Інформаційна політика успішної організації полягає у якісному аналізі і роботи над різнотипними документами. Однак, можливості роботи з кожною інформаційною одиницею, яка стосується організації, немає, оскільки це досить складний процес, який вимагає значних матеріальних та людських ресурсів. Тому, доцільним є свого роду групування цієї інформації і подальша аналітична чи статистична робота над групами. Іншими словами, структуризація цих різнотипних даних.

Дані, що описують реальні об'єкти, процеси і явища, можуть бути представлені в різноманітних формах і мати різний тип та вид.

Дані – це відомості, які характеризують систему, явища, процес чи об'єкт, представлені у визначеній формі і призначені для подальшого їх використання.

За ступенем структурованості виділяються наступні форми представлення даних:

- неструктуровані;
- структуровані;
- слабо структуровані (напівструктуровані).

У сучасному світі досить часто зустрічається поняття «структурована інформація». Однак, не існує чіткого визначення, або переліку критеріїв, які б визначали до якого ж типу віднести ті чи інші дані: структурованих чи неструктурованих. Більше того, у деякій літературі, наукових виданнях та інших друкованих та цифрових джерелах зустрічається термін «напівструктурована» або «слабоструктурована» інформація.

Попри все, в корені цих понять є термін «структура», який являє собою характеристику складу та просторову картину складу об'єкта, взаєморозміщення частин, деталей, елементів, певний взаємозв'язок складових частин об'єкта, його внутрішню будову [4].

Іншими словами, структура – це елементи, з яких складається об'єкт, який можна розглядати як цілісну одиницю. Коли мова йде про неструктуровану інформацію, мається на увазі набір об'єктів (інформаційних одиниць), які між собою не взаємодіють, нічим не пов'язані, і відповідно їх сукупність неможливо розглядати як цілісний об'єкт, а отже, неможливий аналіз над ним. До неструктурованих відносяться дані, довільної форми, включаючи тексти і графіку, мультимедіа (відео, аудіо, фото). Ця форма представлення даних широко використовується, наприклад, в Інтернеті, а самі дані представлені користувачу у вигляді відклику пошукової системи.

Структуровані дані відображають окремі факти предметної області. Структурованими називаються дані, певним чином впорядковані та організовані з ціллю забезпечення можливості застосування до них певних дій (наприклад, візуального або машинного аналізу). Це основна форма представлення відомостей в базах даних.

Організація того чи іншого виду збереження даних (структурованих чи неструктурованих) пов'язана із забезпеченням доступу до них. Під доступом розуміється можливість виділення даних (або множини елементів) серед інших елементів за якимось ознаками із ціллю виконання певних дій над ним або їх групами.

Слабоструктуровані дані – це дані, в яких визначенні певні правила і формати, але у найзагальнішому виді. Наприклад, рядок з адресом, рядок в прайс-листі, П.І.П і т.д. На відмінно від неструктурованих, такі дані з меншим зусиллям перетворюються в структуровану форму, але без процедури трансформації вони також не придатні для аналізу [5].

В основному, більшість методів аналізу даних працює лише з добре структурованою інформацією. Ступінь якості структури впливає на результат, отриманий під час роботи над даними.

Таким чином, документи, як інформаційні одиниці, можуть відноситися до трьох типів даних, що відображено в таблиці 1.

Документ, як інформаційна одиниця, може представлятися у багатьох формах і типах. Такі дані містять в собі безліч характеристик, які потрібні для структуризації документів.

Певному конкретному відділу, який працює з набором документів, або аналітику, для роботи достатньо визначений набір характеристик об'єктів, за якими буде здійснюватися структуризація. Однак, для іншого відділу тієї ж організації, структура і взаємозв'язки між елементами будуть зовсім іншими. Таким чином, в межах однієї організації для роботи над одними і тими ж документами є необхідність в декількох структурах.

Таблиця 1

Типи даних за ступенем структуризації

ДОКУМЕНТ		
Структуровані дані	Напівструктуровані дані	Неструктуровані дані
<ul style="list-style-type: none"> • Поділ інформації на групи та підгрупи; • Групи та підгрупи логічно пов'язані між собою; • Відомості про дані; • Легкі для сприйняття та обробки; • Містять метадані: <ul style="list-style-type: none"> ✓ автор; ✓ зміст документу; ✓ тип; ✓ дата створення; ✓ відповідальні особи; ✓ джерело; ✓ предмет; ✓ опис; ✓ тощо. 		Електронні листи Електронні документи PDF-файли Електронні таблиці Цифрові зображення Відео Аудіо Записи соціальних мереж

Зазвичай створюється універсальна структура, в основі якої лежать певні критерії, однак вона може не повністю задовольняти усі підрозділи організації. Тому, для вибору критеріїв структуризації потрібно попередньо проаналізувати специфіку роботи конкретного закладу, здійснити статистичні обрахунки частоти використання того чи іншого параметру, і на основі цього будувати структуру системи документообігу.

Сучасна організація має справу з динамічними даними. З'являються нові, скажімо як записи в соціальних мережах, які можуть впливати на імідж установи, інші ж стають не актуальними. Окрім нових типів даних, змінюються і критерії за якими документ структуризується. Тому, системи електронного документообігу повинні бути гнучкими, такими, що дозволяють легко включати в їх структуру нові типи даних, не порушуючи всю систему, додавати або забирати критерії структури, таким чином роблячи систему зручною та актуальною в кожен момент часу.

Висновки

Отже, будь-яке сучасне підприємство чи організація працює з великими потоками вхідних та вихідних даних. Їх можна розглядати як інформаційні одиниці (документи), що виконують різні функції.

Окрім документів, що мають юридичну силу, існує ряд інших, не менш важливих в сучасних умовах. Однак, вони часто не розглядаються як інформаційні одиниці, мають обмежену структуру, тобто слабо структуровані, або неструктуровані взагалі. Робота з такими документами є малоефективною, оскільки через відсутність структури не можливий груповий машинний аналіз над цими документами.

Також, з такими документами значно важче працювати людині. Із-за відсутності структури неможлива робота із групами схожих за властивостями документів. Таким чином, структуризація документів дозволяє значно ефективніше вести роботу над ними, скорочувати час на пошук та їх аналіз.

Список використаних джерел

1. Електронний документообіг: сучасні тенденції та проблеми впровадження [Електронний ресурс]. – Режим доступу: http://www.rusnauka.com/34_VPEK_2012/Philologia/7_121024.doc.htm.
2. Варламова Л. Н. Новий стандарт у сфері діловодства / Л. Н. Варламова // Секретар-референт. – 2007. – №6. – с. 164.
3. Российские системы автоматизации документооборота [Електронний ресурс]. – Режим доступу: <http://compress.ru/article.aspx?id=11404>
4. Pullan, Wendy (2000). Structure. Cambridge: CambridgeUniversityPress. ISBN 0-521-78258-9., Структура // Великий тлумачний словник сучасної української мови. - 5-те вид. - К. ; Ірпінь : Перун, 2005. - ISBN 966-569-013-2.)
5. Паклин Н. Бизнес-аналитика: от данных к знаниям / Паклин Н., Орешков В. – Питер, 2013. – 704 с.