

ОБЕКТНО-ОРИЕНТИРАН ПОДХОД ЗА СТРУКТУРИРАНЕ НА СЪДЪРЖАНИЕ

Иван Белев

Главен асистент, доктор, катедра „Информационни технологии и комуникации“, УНСС,
e-mail: i.belev@unwe.bg

Резюме

Научното изследване разглежда обектно-ориентиран подход при структуриране, съхранение и управление на съдържание в организацията. Представени са сценарии, чрез които се прави опит да се демонстрират ползите при използване на обектно-ориентирания подход в системи от тип Enterprise Content Management.

Ключови думи: системи за управление на съдържанието, enterprise content management, ECM, обектно-ориентиран подход, структуриране.

OBJECT-ORIENTED ECM APPROACH

Ivan Belev

Abstract

The paper explores the possibility to use an object-oriented approach for structuring, storing and managing content in the organization. The author presents use cases that try to demonstrate the advantages of using an object-oriented approach in Enterprise Content Management systems.

Keywords: enterprise content management, ECM, object-oriented approach, structuring.

Съдържание и управление на съдържание в организациите

Понятието съдържание заема централна част в контекста на текущата разработка. В дигиталната ера, в която живеем, съдържанието е основният ресурс на голяма част от организациите. То е основа за взимане на решения, както и фактор за иновации и взаимодействие в организациите. Обемът и разнообразието на съдържание, което се създава в организациите или е необходимо за дейността им, достига невиджани досега нива.

В контекста на организациите, съдържание обикновено означава информация във някаква форма. Традиционно съдържанието е под формата на хартиени документи. Но с еволюцията на технологиите и технологичният напредък на организациите, под съдържание вече може да се има предвид и следната информация:

- дигитални документи;
- мултимедийно съдържание;
- съдържание в социални мрежи;
- електронни съобщения;
- съдържание в web страници.

Разглеждайки класификацията по-горе става ясно, че под термина съдържание в общия случай може да се има предвид неструктурирани източници на информация в организацията. Независимо от формата на съдържанието, правилното управление на това богатство от

информация е ключово за успеха на всяка една организация. В това отношение основна роля имат системите за управление на съдържанието в организацията – Enterprise Content Management (ECM). По отношение на понятието Enterprise Content Management, може да приемем българския термин – Управление на съдържанието на предприятието. Авторите Alalwan и Weistroffer [2] правят обобщение на различни определения за ECM от няколко литературни източника ([4], [5] и [6]). общите точки в обобщените от тях дефиниции определят термина Управление на съдържанието на предприятието като съвкупност от стратегии, методи, инструменти, процеси, умения и технологии за управление на пълния жизнен цикъл на съдържанието в организацията – създаване/извличане, откриване, съхраняване, управление, опазване, търсене/намиране, публикуване и предоставяне. Gartner [7] добавят към определението за Enterprise Content Management и анализа на използването на съдържанието в организацията с цел по-добро осигуряване на правилното съдържание на крайните ползватели, където те имат нужда от него.

Традиционен подход за съхранение и структуриране на съдържание

Поглеждайки назад във времето, съхранението и структурирането на съдържание е много различно от това, което технологиите за управление на съдържанието предоставят. Преди използването на софтуер за целите на управление на съдържанието, организациите съхраняват своите документи, използвайки:

- физически папки.
- кутии.
- шкафове.
- библиотеки.
- складови помещения.

За откриването и управлението на документите се създават и поддържат индекси, каталози и други форми на обозначение, чрез които служител собственоръчно открива необходимия документ.

С въвеждането и разпространението на компютърните технологии, съхранението и структурирането на съдържание в организациите се премества от физически носители към дигитални форми на съдържание. Организациите създават дигитално пространство, в което съхраняват дигиталните си документи и друго съдържание. За да бъде лесно откриваемо се прилага дигитална аналогия на физическите папки – компютърни директории. Използвайки директории и възможността те да бъдат именувани по избран начин, както и тяхното свойство да могат да помещават други директории, образувайки йерархия, дава възможност на организациите да приложат подход за структуриране на съдържанието, с който да осигурят дигитален аналог на описаното в предходния абзац – форма на обозначение, чрез която да се осигури по-лесно откриване на файлове/документи.

Прилагайки описания по-горе подход на практика означава следното – организациите създават йерархична структура от компютърни директории, организирано по специфичен за организацията начин, в която разпределят дигиталните документи/файлове. При необходимост за търсене на даден документ/файл в последствие, той бива намерен като се търси в съответната директория в структурата. За целите на научното изследване ще бъде представен практически пример, чрез който да се илюстрират разглежданите подходи за структуриране и техните предимства и недостатъци. Представяйки различните подходи за структуриране по-долу и практическите примери, авторът ще използва термина „документи“ като обобщаващ термин за различните видове съдържание в дигитален формат.

Организация може да започне да съхранява своите документи/файлове в директории, като организира на първо ниво отделни директории за клиентските документи и за други вътрешни

за организацията документи, които не са свързани с нейните клиенти. Разглеждайки клиентските документи, на второ ниво може да се създаде отделна директория за всеки един клиент, в която да се съхраняват документите на дадения клиент. В случай, че клиентите имат дългосрочна история в организацията, на следващо ниво може да се създаде директория за всяка отделна година, в която организацията съхранява документи за дадения клиент. Всяка отделна директория на ниво година. В случай, че документите в рамките на един клиент за една година са много, на следващо ниво може да се създаде различна директория за основни типове документи, които са обект на търсене. Ако погледнем от гледна точка на търговски организации, това може да бъдат отделни директории някои често срещани типове документи в тази сфера:

- Оферти.
- Договори.
- Фактури.
- и др.

При така представения примерен сценарий, оферта на даден клиент (Client A) от 2024 година ще бъде съхранена в следната директория:

Клиентски документи -> Client A -> 2024 -> Оферти



Фигура 1: Пример за традиционен подход за съхранение и структуриране на съдържание в йерархия от директории

Този подход за структуриране и съхранение на съдържание би работил без особени проблеми при сравнително статични параметри, които се използват в организацията за организиране и откриване на съдържание, както и при сравнително малък брой уникални стойности на различните нива в йерархията.

В посочения пример става ясно, че ключова роля за структуриране на съдържанието в организацията и в последствие неговото лесно откриване имат именно параметрите или данните за всеки един съхраняван документ като единица съдържание. Тези параметри в терминологията на Enterprise Content Management се наричат свойства на документа (properties) или също така – метаданни (metadata). Милев и Табов [3] дефинират метаданните като структурирана информация, която описва, обяснява, открива или дава възможност лесно да бъде открит даден информационен ресурс.

Така реализирания подход за съхранение и структуриране на съдържание е приложим в организациите до достигане на определени нива на динамика в дейността, както и до достигане на определени обеми от съдържание. Бързо развиващата се среда на дейност на организациите от всякакъв тип в днешните условия водят до повишени нива на динамика и големи обеми, които извеждат на преден план недостатъците на представения традиционен подход:

- трудна навигация в структурата от директории при голям брой директории на всяко ниво от йерархията.
- трудно управление на структурата при постоянно променящи се изисквания по отношение на параметрите (метаданните) на отделните единици съдържание.
- трудно извличане на документи при необходимост да се търси по повече от една стойност на метаданните.

Може да бъде представен следните примери, за да се илюстрират недостатъците на този подход за структуриране и съхранение на съдържанието:

- При голям брой клиенти с дълга история в организацията (например в порядъка на милиони) би било много трудно да се открие директорията на конкретния клиент. Освен това всяко търсене трябва да премине през последователно намиране на съответната директория на всяко едно от нивата в йерархичната структура.
- При поява на евентуално ново изискване за отделяне на клиентите по техния тип например (физически лица и юридически лица, или по сегмент на клиента) ще се наложи структурата да бъде преобразувана като се добавят нови нива и се разпределят текущите директории в тях.
- Евентуално изискване да се намерят и извадят всички фактури за две избрани години без значение кой е клиентът би довело до множество търсения на различни нива, особено при голям брой клиенти. В този случай трябва да се търси вариант за допълнителна организация на документите само по техния тип например.

Описаното дотук дава основание да се потърси друг подход за структуриране на съдържанието. Такъв подход се предлага от системите за управление на съдържание от тип Enterprise Content Management.

Съхранение и структуриране на съдържание в ЕСМ

ЕСМ системите въвеждат допълнителен компонент за да съхраняват и използват в структуриран вид метаданните на съдържанието в организацията – обикновено чрез база от данни. Всички метаданни за даден документ (единица съдържание) се съхраняват като запис в база от данни заедно с описание на „пътя“ до физическото разположение на документа/файла. Документите продължават да бъдат съхранявани в структура от директории, но тя вече не се използва за йерархия, в която да се търси по името на определената директория.

Следвайки примера, който бе използван до момента, след въвеждане на описаната по-горе организация за съхраняване на документите, намирането на фактурите на клиент Client A за 2024-та година би било възможно след следната заявка в базата данни:

Всички документи за Име на клиент: Client A, Година: 2024, Тип на документ: Оферта

За разлика от традиционния подход, при този подход лесно може да бъдат правени сечения и комбинации между стойностите на различните метаданни, например:

Всички документи за Година: 2024, Тип на документ: Оферта или Договор

Използвайки такъв подход решава голяма част от проблемите описани до момента, но някои от тях остават, като се прехвърлят на друго ниво – в организацията на структурата на базата данни. Повишената динамика на изискванията по отношение на метаданните на документите и техните различни стойности води до трудности по отношение на:

- броя колони в базата данни.
- номенклатурата от стойности в базата данни.
- обособяването на различни таблици в базата данни за съхранение на документи с различен набор от метаданни.

Може да се представи следният опростен пример, за да се илюстрират част от описаните проблеми: Ако в организацията е необходимо да се съхраняват общи клиентски документи, за които има един набор от метаданни:

- тип на клиент (например физическо и юридическо лице).
- идентификатор на клиент.
- категория на документ (например Документи за идентификация).
- тип на документ (например лична карта, паспорт, пълномощно).
- дата на документ.
- потребител.

В същото време е необходимо да се съхраняват вътрешни документи с правно естество, които не са свързани с определен клиент и имат друг набор от метаданни:

- номер на дело.
- отдел/департамент.
- тип на документ.
- дата на документ.
- срок за резолюция.

Може да се добави и още един пример – съхраняване на документи, свързани с продуктите на клиента, при които има следния набор от метаданни:

- тип на клиент (например физическо и юридическо лице).
- идентификатор на клиент.
- категория на документ (например Договорни документи).
- тип на документ (например договор, анекс, фактура).
- Тип на продукт (например застраховка, кредит)
- дата на документ.
- потребител.

Разглеждайки изискванията за структуриране на съдържанието в посочените три примера се вижда, че в структурата на базата от данни за съхранение на метаданните на документите е необходимо да има различни таблици, в които да се разделят отделните видове документи. Същевременно между тях остават някои общи метаданни, които ще бъдат дублирани между отделните таблици. Това би довело до проблеми при извличането на данни, подобни на проблемите с традиционния подход. Може да се наложи да се обединяват резултати от повече от една заявка към базата от данни. Също така е възможно да има трудности при дефиниране и управление на колони за метаданни, които присъстват едновременно при повече от един вид документи – необходимост от индексирание, обединяване, сортиране, филтриране на резултати.

За решаване на част от тези проблеми може да бъде приложен изменен подход за съхранение и структуриране на съдържание в системите за управление на съдържанието, който е обектно-ориентиран.

Обектно-ориентиран подход за структуриране на съдържание в ЕСМ

Еволюцията на технологии води до дефинирането на обектно-ориентирания подход за реализация на различни задачи в света на ИТ. Може би най-разпознаваем е терминът за обектно-ориентирано програмиране, използван в редица програмни езици и технологии. Стъпвайки на предимствата на подход като дизайн модел [8], някои системи за управление на съдържанието успешно го прилагат за структуриране на метаданните на отделните единици съдържание (документи) в системата. Такава система например е IBM FileNet P8 Platform [9].

Използвайки обектно-ориентиран подход за структуриране на съдържанието в ЕСМ системите може да доведе до елиминиране на недостатъците, описани до момента в изследването. При този подход всички структурно-различни видове съдържание (документи) могат да се дефинират като отделни класове (и под-класове) в хоризонтална и вертикална (йерархична) структура. При създаването на структурата се следват няколко основни правила:

- Всички метаданни се дефинират като отделни *properties*, които могат да бъдат използвани в един или повече класове документи.
- Всеки клас документ може да приема определен набор от избрани метаданни (*properties*).
- Всеки под-клас наследява метаданните на своя „родител“ (*parent class*) и от своя страна може да приема и други метаданни, специфични за него, както и да има свои под-класове.

За представяне на предимствата на обектно-ориентирания подход за структуриране на съдържанието може отново да се даде пример, който е сходен на използваните до момента в изследването:

На най-високо ниво се дефинира клас Документ със следните метаданни:

- Заглавие.
- Дата на създаване.
- Потребител.

На следващо ниво като под-клас на клас Документ се дефинира клас Документ на организацията, който наследява метаданните на клас Документ и има в допълнение следните метаданни:

- Тип на документ.
- Статус.

На следващо ниво се създават два под-класа – Клиентски документ и Документ Човешки ресурси, които наследява метаданните на клас Документ и клас Документ на организацията, но в допълнение имат и собствени метаданни. Например за клас Клиентски документ:

- Идентификатор на клиент.
- Тип на клиент.

На последно ниво се дефинира клас Продуктов документ, който е под-клас на Клиентски документ, наследява всички метаданни нагоре в йерархията и има следните допълнителни метаданни:

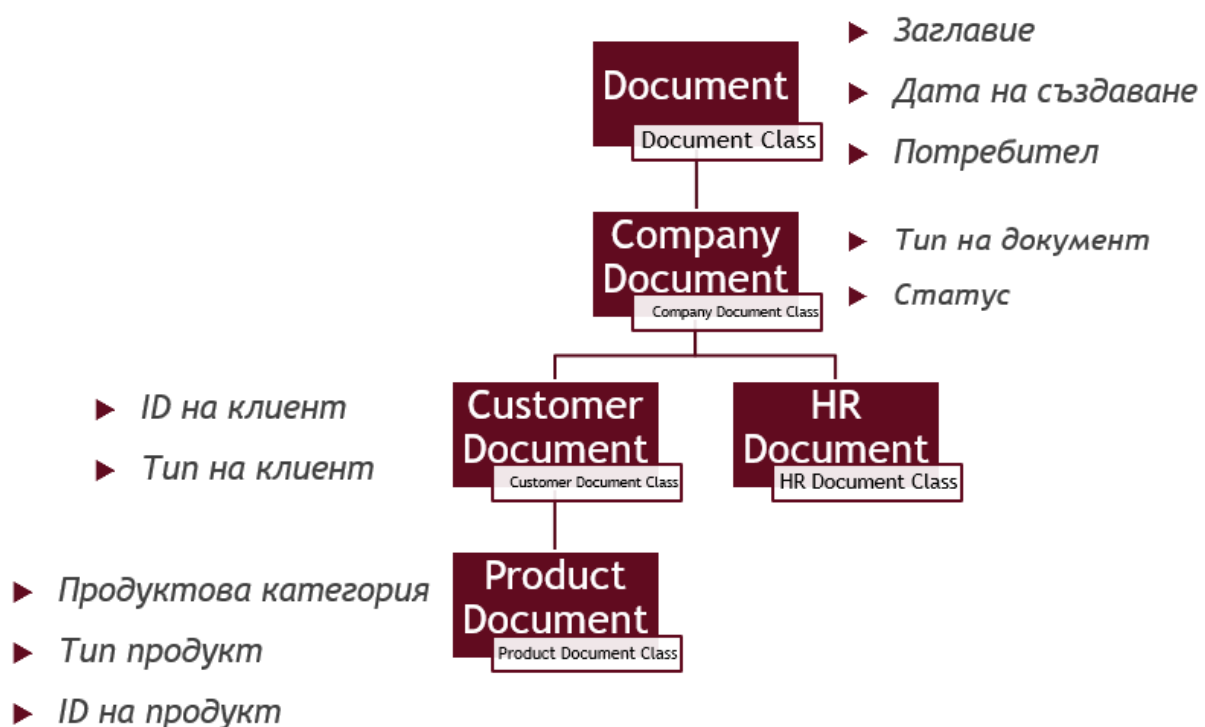
- Продуктова категория.
- Тип продукт.
- Идентификатор на продукт.

Така дефинираната структура дава следните предимства:

- На лице е ясно разграничение между структурно различните видове документи в организацията.

- Могат да се търсят различните видове документи, като се използва критерия за клас на документа в комбинация с метаданните на конкретния клас, независимо дали са негови собствени или наследени от „родителските“ класове.
- Същевременно е възможно да се търсят документи от даден клас, като търсенето да включва и документи от неговите под-класове, които отговарят на зададените критерии.
- Структурата се управлява сравнително лесно като за всеки нов клас или под-клас, може да се използват отново предварително дефинираните метаданни (properties) без да е необходимо да се създават нови. Това позволява лесно да се търси по един и същ критерий между различни класове и под-класове, които притежават общи метаданни.
- Сравнително лесно управление на структурата чрез добавяне на нови метаданни на високо ниво в даден клас, което води до автоматичното им създаване и в техните под-класове.

Дефинираният пример може да бъде представен визуално чрез следната диаграма:



Фигура 2: Пример за обектно-ориентиран подход за съхранение и структуриране на съдържание

Реализирайки подобен подход за структуриране на съдържанието в организацията дава възможност да се използват предимствата на системите от Enterprise Content Management по отношение на бързото намиране на дадена единица съдържание в комбинация с елиминиране на трудностите, свързани с дефиниране и управление на самата структура.

Могат да бъдат представени следните примери, за заявки за извличане на съдържание в дефинираната примерна обектно-ориентирана структура:

Всички документи от клас Продуктов документ за ID на клиент: Client A, Тип на продукт: Застраховка, Потребител: Иван Иванов, Дата на създаване: през 2023 година.

Всички документи от клас Клиентски документ (включително документи от под-клас Продуктови документи) за Тип на клиент: Физическо лице, Тип на документ: Фактура, Статус на документ: Активен, Дата на документ: през месец Октомври 2023

Заклучение

Динамичният свят, в който всички организации развиват своята дейност поставя редица предизвикателства пред тях по отношение на съдържанието в организацията и управлението на неговия жизнен цикъл. За справяне с предизвикателствата организациите могат да използват системи за управление на съдържанието от тип Enterprise Content Management, избирайки такава система, която дава възможност да се приложи обектно-ориентиран подход при създаването на структурата на съдържанието. Използвайки такъв подход организациите могат максимално да се възползват от предимствата му и да успеят да се справят в голяма степен с трудностите, свързани със структурирането и управлението на съдържанието в организацията.

References

1. Cameron, S. A. (2011). *Enterprise content management: a business and technical guide*. BCS, The Chartered Institute.
2. Alalwan, J. A., & Weistroffer, H. R. (2012). *Enterprise content management research: a comprehensive review*. *Journal of Enterprise Information Management*, 25(5), 441-461.
3. Milev, P., & Tabov, Y. (2022). *Conceptual Approach For Presenting Text Data From Web-Based Information Systems In Structured Form*.
4. Smith, H.A. and McKeen, J.D. (2003), “*Developments in practice VIII: enterprise content management*”, *Communications of the Association for Information Systems*, Vol. 11, pp. 647-59
5. Woolley, R. and Fletcher, D. (2007), “*Research summary: enterprise content management*”, white paper, Department of Technology Services, available at: <http://utah.pdfs.com/awweb/awarchive?type=file&item=18380>
6. Fowler, D. (2008), “*Implementing enterprise content management using Microsoft SharePoint*”, Master’s thesis, University of Oregon, Eugene, OR
7. *Definition of Enterprise Content Management (ECM)* - Gartner Information Technology Glossary. (n.d.). Gartner. <https://www.gartner.com/en/information-technology/glossary/enterprise-content-management-ecm>
8. Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design patterns: Elements of reusable object-oriented software*. Addison Wesley.
9. Zhu, W.-D., Buchanan, N., Oland, M., Poggensee, T., Romero, P. E., Snow, C., Worel, M., & Redbooks, I. B. M. (2011). *IBM FileNet P8 Platform and architecture*. IBM