

Jadwiga Sobieska-Karpińska, Marcin Żmigrodzki

STRUKTURA INFORMACJI REPREZENTUJĄCEJ WIEDZĘ

1. Wstęp

Według raportu Forrester Research rynek rozwiązań wspierających zarządzanie informacją słabo ustrukturalizowaną w 2003 roku wyniósł 1,6 mld dol., a w roku 2006 ma wynieść 3,3 mld dol. [Kontzer 2004]. Wśród głównych „graczy” można tutaj wymienić IBM, Open Text, Documentum, Interwoven czy Vignette. Niektórzy autorzy zaczęli zauważać, że ten rynek zaczyna zlewać się z dojrzałym rynkiem rozwiązań wspierających zarządzanie informacją silnie ustrukturalizowaną (np.: hurtownie danych, analizy statystyczne, eksploracja danych, business intelligence) [Miller 2003].

Wobec rosnącego zainteresowania zarządzaniem informacją słabo ustrukturalizowaną warto zdefiniować, czym jest struktura takiej informacji i jakie jest jej znaczenie w przedsiębiorstwie.

W poniższym artykule zdefiniowano strukturę informacji reprezentującej wiedzę na gruncie infologii oraz scharakteryzowano jej rolę w zarządzaniu wiedzą. Następnie przedstawiono proces strukturalizacji takiej informacji, którego celem jest wzmocnienie jej struktury. Założono też, że silniejsza struktura pozwala lepiej zarządzać taką informacją, ponieważ łatwiej ją zlokalizować w bazie wiedzy, utrzymać jej jakość w czasie oraz zinterpretować i przyswoić przez odbiorców. W końcu jako przykład przedstawiono informacje z pewnego systemu zarządzania wiedzą o różnej sile ustrukturalizowania.

Według literatury z zakresu infologii [Stefanowicz 2004] informacja (I) jest funkcją komunikatu (K). W wyidealizowanym przypadku może być niezależna od odbiorcy i kontekstu (stosuje się tu określenie informacji na poziomie datalogicznym) i zapisuje się wzorem:

$$I(K). \quad (1)$$

Taka definicja informacji jest bliska definicji terminu dane według rozumienia literatury zarządzania wiedzą [Devenport, Prusak 1998].

Z punktu widzenia informacji reprezentującej wiedzę trafniejsze jest ujęcie informacji, w którym odbiorca przyswaja sobie fragment rzeczywistości i tworzy

jego odwzorowanie w umyśle (poziom infologiczny). To odwzorowanie jest informacją o danym wycinku rzeczywistości. Efektem odbioru rzeczywistości jest wyodrębnienie w nim obiektów, które mają określone cechy i pomiędzy którymi zachodzą relacje.

W tym wypadku informacja opisana jest funkcją komunikatu K , użytkownika U , który ją odbiera i zadania Q , w kontekście, którego jest odbierana. Prezentuje to poniższy wzór:

$$I(K, U, Q). \quad (2)$$

Taka informacja jest subiektywna w odróżnieniu od tej określonej wzorem (1). Można ją interpretować w ten sposób, że na informację (I), którą odbiera użytkownik składa się komunikat (K), jego sposób odbioru (U) oraz zadanie, które na podstawie tej informacji musi wykonać (Q). Subiektywność odbioru informacji zależy od szeregu czynników, takich jak [Stefanowicz 2004]:

- czasu do namysłu – czy decyzja, która ma być podjęta na podstawie informacji jest pilna czy nie;
- dotychczas posiadanej wiedzy – w kontekście zarządzania wiedzą jest to najważniejszy czynnik. Różnice wynikają z tego, że w zależności od posiadanej wiedzy można wyciągnąć różne wnioski z danej informacji;
- problemu – cel użycia informacji powoduje selektywne przyswajanie informacji;
- stanu emocjonalnego użytkownika – wiele źródeł z zakresu psychologii kognitywnej, dydaktyki wskazuje, że nastawienie emocjonalne użytkownika ma duży wpływ na to, jak i w jakim zakresie przyswaja informację.

Sam komunikat K zdefiniowany jest poniższym układem:

$$K := \langle O, P, t, v \rangle, \quad (3)$$

gdzie: „ O – obiekt należący do analizowanej rzeczywistości R ; przy tym O może oznaczać dowolny obiekt materialny, proces, zdarzenie, pojęcie abstrakcyjne, własność innego obiektu itd. – wszystko, co można nazwać w języku naturalnym za pomocą rzeczownika (...);

P – predykat określający wartość cechy A obiektu O lub jego związek z innymi obiektami należącymi także do R ;

t – czas, w którym obiekt O jest rozpatrywany ze względu na P (...);

v – wektor szczególnych warunków związanych z obiektem O , predykatem P i czasem t ” [Stefanowicz 2004, s. 15].

Jak podkreśla się w literaturze przedmiotu [Stefanowicz 2004], predykat P może odnosić się do cech samego obiektu O lub relacji danego obiektu O do innego O' . Na potrzeby rozważań o strukturze informacji reprezentującej wiedzę można pominąć drugą z tych sytuacji i wziąć pod uwagę tylko pojęcie struktury samej informacji. W takim razie wzór (3) przybierze poniższą postać:

$$K := \langle O, A = a, t, v \rangle, \quad (4)$$

gdzie: O – rozpatrywany obiekt,
 A – atrybut (cecha) obiektu O ,
 a – wartość atrybutu A ,
 t – czas, w którym obiekt O ma wartość a z atrybutu A ,
 v – wektor dodatkowych charakterystyk związanych z obiektem O ,
 atrybutem A i jego wartością a oraz czasem t .

Komunikat K opisany wzorem (4) można interpretować tak, że obiekt O ma cechę A o wartości a w czasie t i w okolicznościach v .

Atrybut A może pełnić różne funkcje w zależności rodzaju informacji, jaką reprezentuje komunikat K , a mianowicie:

- może opisywać stan obiektu O (informacje faktograficzna, komparatywna, semantyczna, klasyfikacyjna, przestrzenna, kwerencyjna);
- może być strukturą obiektu O (informacja strukturalna);
- może być procedurą postępowania z obiektem O (informacje proceduralna, normatywna, imperatywna).

Z punktu widzenia tematu niniejszego artykułu brana pod uwagę jest tylko i wyłącznie sytuacja, gdy A oznacza strukturę obiektu O (informacja strukturalna).

W niniejszej publikacji rozszerzono rozważania zawarte w [Stefanowicz 2004] i przyjęto, że strukturę informacji można rozumieć jako komunikat K' , określający, jak interpretować czy rozumieć pewne elementy informacji, jaką niesie komunikat K (dalej nazywany treścią), do którego struktura się odnosi. Struktura zatem determinuje i ogranicza znaczenie informacji na niższym poziomie. Struktura K' jest określona wzorem (4):

$$K' := \langle O', A' = a', t', v' \rangle.$$

Struktura jest pojęciem znajdującym się na wyższym poziomie w stosunku do treści. Inaczej mówiąc, struktura odnosi się do treści, jest opisem treści. Ujmując problem odwrotnie, można powiedzieć, że sposoby interpretacji treści są ograniczone przez strukturę.

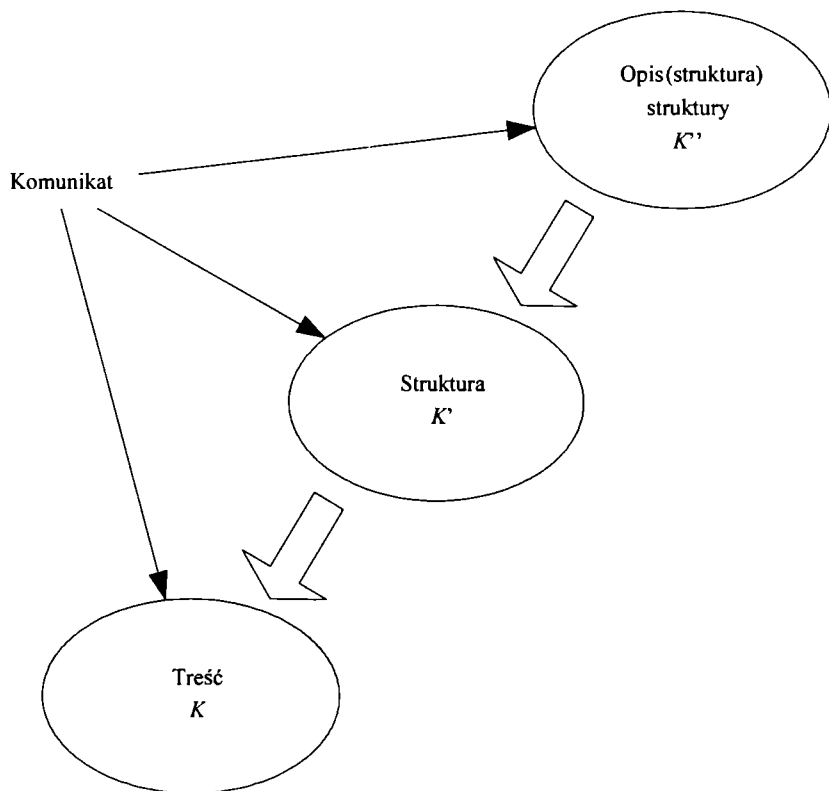
Skoro sama struktura jest również komunikatem, to może być podporządkowana innej strukturze. Tę zależność metaforycznie przedstawiono na rys. 1.

Dla uproszczenia rozważań pomija się kwestie sposobu odbioru informacji przez użytkownika U i kontekstu zadania Q (por. (2)). Należy jedynie powiedzieć, że wpływają one na interpretację komunikatu, a przez to na sposób tworzenia jego struktury. Dalej mówi się tylko o strukturalizacji **komunikatu**.

Na rys. 1 przedstawiono wielopoziomową naturę komunikatów w kontekście ich struktury. Wielopoziomowość wynika z tego, że opis wyższego rzędu (struktura) jest również komunikatem, który może być w określony sposób ustrukturalizowany.

Komunikat wyższego rzędu niekiedy określa się mianem metainformacji. „Meta informację należy interpretować jako treść komunikatu K , w którym obiektem jest inna informacja $I(K')$ zawarta w innym komunikacie K' , bądź jakiś

element rozpatrywanego systemu informacyjnego: jego zasoby informacyjne, źródła informacji, jego użytkownicy, zasady korzystania z systemu, usługi informacyjne świadczone przez system i inne. Jest to więc pojęcie względne: to, co dla jednego użytkownika jest metainformacją, dla innego może być zwykłą informacją” [Stefanowicz 2004, s. 48].



Rys. 1. Poziomy struktury komunikatów

Źródło: opracowanie własne.

Struktura może pełnić następujące funkcje:

- 1) informacyjną – poprzez opisywanie innych informacji (zasobów informacyjnych, infrastruktury, sposobu korzystania z systemu informacyjnego itp.) sama informuje, zapewniając lepszą obsługę tychże;
- 2) koordynacyjną – umożliwia koordynację wielu procesów przetwarzania różnych informacji w celu osiągnięcia globalnego rezultatu;
- 3) kontrolną – pozwala na weryfikację jakości informacji poprzez wskazanie i możliwość odniesienia się do standardów zapisanych w metainformacjach;
- 4) usługową – wymienia się tutaj informacje o użytkownikach systemu wykorzystywane w celu podnoszenia przyjazności systemu (np. poprzez personalizację).

Zatem strukturę komunikatu zdefiniować można jako opis wyższego rzędu odnoszący się do danego komunikatu (treści), który wskazuje na wybrane jego cechy¹.

2. Strukturalizacja informacji

Strukturę komunikatu S w omawianym kontekście można zdefiniować jako zbiór atrybutów, ich dziedzin oraz relacji między nimi, które odnoszą się do właściwej porcji informacji. Budowanie struktury komunikatu określono tutaj mianem strukturalizacji.

Strukturalizację komunikatu definiujemy jako proces przypisywania wartości atrybutów ze zbioru S do komunikatu K . Efektem procesu strukturalizacji jest komunikat ustrukturalizowany K_S . Zatem można zapisać następujące równanie operacji strukturalizacji:

$$K + S = K_S. \quad (5)$$

Komunikat ustrukturalizowany K_S charakteryzuje się tym, że ma jednoznacznie przypisane atrybuty z S do pewnych jego fragmentów. W trakcie tego procesu nie wszystkie atrybuty ze zbioru S muszą uzyskać wartości dla komunikatu K , więc zbiór atrybutów opisujących K jest podzbiorem S . Zatem:

$$K_S = \{A_1 = a_1, \dots, A_m = a_m\},$$

gdzie $m: \leq n$

a_i – wartość atrybutu A_i dla danej porcji informacji.

Trzymając się formalnego zapisu prezentowanego przez B. Stefanowicza, relację między strukturą a treścią można zapisać w poniższy sposób:

$$a' = a,$$

$$\text{dla } K' := p(O', A' = a', t', v'),$$

$$K := p(O, A = a, t, v),$$

$$\text{stąd } K_S := p(O, a' = a, t, v), \quad (6)$$

gdzie: K' – komunikat będący strukturą,

K – komunikat będący treścią,

K_S – komunikat K ustrukturalizowany za pomocą K' .

¹ Podejściu, które zakłada wielopoziomowość informacji, wydaje się być bliska postawa badawcza wywodząca się z nauk humanistycznych, określana jako strukturalizm, „kładąca nacisk na analizę struktury badanych zjawisk, a nie na ich genezę lub funkcję...” [Encyklopedia PWN... 2003, hasło – strukturalizm]. Strukturalizm zakłada istnienie w trwałych, ukrytych struktur, leżących u podłoża badanych zjawisk, które same w sobie niekoniecznie muszą być trwałe. Odkrycie tych struktur możliwe jest przez tworzenie abstrakcyjnych modeli w procesie dedukcji. Podejście strukturalistyczne znalazło zastosowanie w teorii kultury, etnologii, socjologii, językoznawstwie, semiotyce.

W takim ujęciu, jak przedstawiono w równaniu [6], strukturalizacja jest operacją składania dwóch komunikatów K i K' taką, że wartości atrybutów a' traktuje się jak atrybuty A na niższym poziomie komunikatu K i przypisuje się im wartości a .

Jak już wspomniano, atrybut A w trakcie strukturalizacji powinien przyjmować sensowne wartości a , zgodne z jego znaczeniem. W przeciwnym wypadku strukturalizacja nie wygeneruje użytecznej informacji K_S i nie będzie miała praktycznego zastosowania. Zatem warunkiem koniecznym powodzenia strukturalizacji jest **interpretowalność komunikatu ustrukturalizowanego w kontekście struktury**. Inaczej mówiąc, równanie $a' = a$ musi nieść sensowne dla odbiorcy znaczenie. K' zawiera wartości atrybutów interpretowalne jako atrybuty w K . To oznacza, że przy strukturalizacji komunikatu trzeba brać pod uwagę kontekst U i Q (por. (2)).

Przy przyjęciu koncepcji przedstawionej na rys. 1, czyli uznaniu wielopoziomowości struktur (metainformacji) i tego, że one również same są komunikatem, można zapisać, że:

$$S = K$$

oraz $K_S = K$.

Zatem zarówno struktura, jak i informacja ustrukturalizowana mogą być dalej strukturalizowane.

Należy jeszcze wspomnieć o pewnej szczególnej właściwości operacji strukturalizowania komunikatów: **wielokrotne strukturalizowanie tej samej treści za pomocą tej samej struktury nie rozszerza pola znaczeniowego**. Ów fakt wynika z tego, że w ramach tej samej interpretacji kolejne operacje składania K_S działają na tych samych argumentach. Przedstawiono to na poniższych równaniach:

$$K' := p(O', A' = a', t', v'),$$

$$K := p(O, A = a, t, v),$$

$$K_S := p(O, a' = a, t, v),$$

$$K_S^1 := p(O, a' = a^{-1}, t, v),$$

gdzie: K' – struktura,

K – treść,

K_S – komunikat K ustrukturalizowany za pomocą K' ,

K_S^1 – komunikat K_S powtórnie ustrukturalizowany za pomocą K' .

Zatem można stworzyć zależność:

jeżeli

$$\pi(a) \geq \pi(a^{-1}),$$

to

$$\pi(K_S) \geq \pi(K_S^1). \quad (7)$$

Dzieje się tak dlatego, że w kolejnych operacjach strukturalizacji mogą brać udział nie wszystkie atrybuty komunikatu K , a następnie K^1 .

Strukturę komunikatu można podzielić na kilka typów:

1) ukryta – jest efektem aktywności umysłowej ludzi. Tym aspektem zajmuje się nie tylko zarządzanie wiedzą [Nonaka, Takeuchi 2000], ale m.in. psychologia poznawcza [Anderson 1998; Tavis, Wade 1995], rozwojowa [Birch, Malim 1998], dydaktyka. Struktura w tym wypadku może być utożsamiana z modelami myślowymi, pojęciowymi o przedmiocie rozważań. Nie obejmują jej prezentowane tutaj rozważania.

2) jawna – jest zapisana jako odrębna porcja informacji, zwykle w systemie informatycznym. Niekiedy potocznie jest nazywana metadanymi lub metainformacją. Jest sformalizowanym opisem odnoszącym się do danej informacji, który determinuje jej sposób przetwarzania i przyswajania². Strukturę jawną można podzielić na:

a) zewnętrzną – jest zapisana w systemie informatycznym poza samym komunikatem K (treścią). Z reguły jej zadaniem jest umiejscowienie treści w kontekście innych treści przechowywanych w systemie informatycznym. Są to na przykład kategorie, atrybuty, daty itp.,

b) wewnętrzną – jest zapisana w ramach komunikatu K (treści) i dzieli go na części. Zwykle celem struktury wewnętrznej jest uporządkowanie treści i ułatwienie przyswojenia jej. Są to na przykład tytuły, sekcje, nagłówki, wypunktowania, stopki, ramki, akapity itp.

3. Pomiar wielkości struktury

Literatura przedmiotu o pomiarze struktury informacji wspomina przy okazji pomiaru ilości informacji dostarczanej przez uwzględnienie w komunikacie K atrybutu A oraz jego konkretnej wartości a . „... $\mu(A)$ – ilość informacji wynikająca z uwzględnienia w komunikacie K atrybutu A ; ilość ta może być (...) wyliczona według miary Hartleya, gdzie w miejsce n należy wstawić liczbę wszystkich atrybutów należących do zbioru zawierającego atrybut A ; $\mu(a)$ – ilość informacji zależna od wartości, jaką przyjmuje atrybut A . (...) Jeżeli atrybut A oznacza strukturę obiektu O , a zaś jest opisem konkretnej struktury, to $\mu(a)$ może być wyliczone na podstawie miary Hartleya. Jeśli jest to struktura złożona, o znanych wewnętrznych powiązaniach, to wzór [Hartleya – przyp. autorów] powinien być tak zmodyfikowany, aby można było uwzględnić te zależności. Modyfikacja może polegać na przypisaniu każdemu elementowi tej struktury określonej wagi, odzwierciedlającej merytoryczną więź tego elementu w całej strukturze” [Stefanowicz 2004, s. 87]. Wzór Hartleya przybiera postać:

² Jeżeli dalej używa się terminu „struktura”, to należy pamiętać, że odnosi się on w domyśle do struktury jawnej.

$$\mu(n) = \log n, \quad (8)$$

gdzie: $\mu(n)$ – miara ilości informacji n ,
 n – liczba elementów w pewnym zbiorze obiektów, które reprezentują mierzoną informację.

Gdy logarytm ze wzoru (8) ma podstawę 2, jednostką miary są bity. Jest to jednostka ogólnie zaakceptowana przez normy ISO i taka jest przyjęta w dalszych rozważaniach.

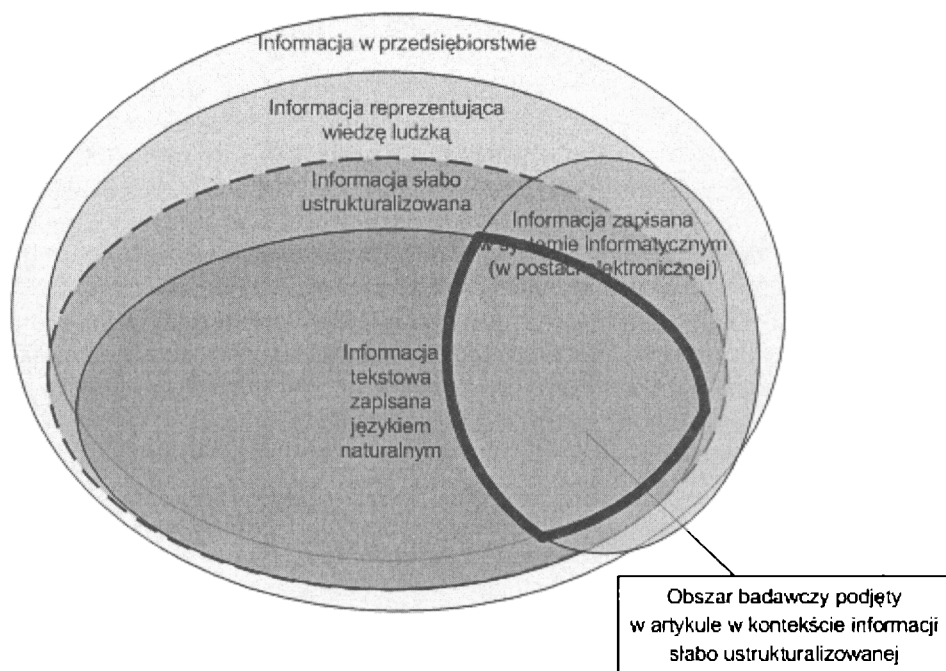
Wielkość struktury określana może być przez minimalną ilość informacji potrzebnej do zapisania danej struktury, czyli wielkość opisu jeszcze wyższego rzędu K'' , wymaganego do opisanie opisu odnoszącego się do treści. Nawiązując do wcześniej przytoczonych równań, proponuje się przyjąć, że wielkość struktury K' to logarytm o podstawie 2 z liczby atrybutów ze zbioru A (czyli a' , bowiem dla $K_S A$ jest równe a') użytych do opisanie komunikatu K_S .

Taki sposób pomiaru powinien jeszcze spełniać warunek interpretowalności a' w kontekście K' , więc może zostać użyty tylko do pomiaru względnego, czyli porównania wielkości struktur dwóch komunikatów względem siebie. Aby zmierzyć względną wielkość, trzeba wcześniej sprowadzić porównywane struktury do wspólnego mianownika, czyli zapisać je za pomocą takiej samej struktury jeszcze wyższego rzędu K'' (najwyższa elipsa na rys. 1) po to, aby spełnić warunek interpretowalności informacji.

4. Informacja reprezentująca wiedzę

Na rysunku 2 schematycznie przedstawiono relacje między różnymi rodzajami zasobów informacji w przedsiębiorstwie.

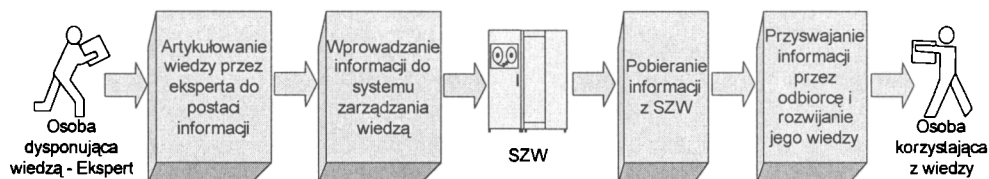
O dużej części informacji obecnej w przedsiębiorstwie można powiedzieć, że reprezentuje wiedzę ludzką. Z tego część informacji, jak podają źródła ok. 80-90% [Langham 2003; Kugel 2003; Trombly 2003], jest słabo ustrukturalizowana, z czego zdecydowana większość ma postać tekstów pisanych językiem naturalnym [Jones 2000]. Zgodnie zaś z badaniami Gartnera od 80 do 100% informacji rocznie jest konwertowane na postać elektroniczną [Trombly 2003] przez skanowanie lub ewentualnie przepisywanie. Z kolei tylko pewna część informacji jest przechowywana w postaci elektronicznej w systemach informatycznych. Zatem niniejsze rozważania dotyczą pogrubionego obszaru na rys. 2. Powszechnie do określenia tego rodzaju informacji używa się terminu informacja nieustrukturalizowana [Ark Group...; Devenport, Prusak 1998; Ferruci, Lally 2004; *Knowledge Management in the...* 2001; *Knowledge Management Magazine*; Langham 2003; Pai, Miller, Honovar, Wong 2002; Smith, Farquhar 2000; Stensmo, Thorson 2003]. Informacja silnie ustrukturalizowana to otoczenie zbioru informacji słabo ustrukturalizowanej (zaznaczonego linią przerywaną na rys. 2).



Rys. 2. Schematyczne umiejscowienie informacji słabo ustrukturalizowanej reprezentującej wiedzę w informacji w przedsiębiorstwie

Źródło: opracowanie własne.

Rysunek 3 prezentuje schematycznie konwersję między wiedzą i informacją w komunikacji wspomaganą systemem zarządzania wiedzą (SZW).



Rys. 3. Schemat konwersji wiedzy w informację i z powrotem w wiedzę za pośrednictwem systemu zarządzania wiedzą

Źródło: opracowanie własne.

Informacja, która reprezentuje wiedzę, przybiera z reguły postać tekstu zapisanego językiem naturalnym. Jest to dosyć oczywiste m.in. dlatego, że większość zapisanej ludzkiej wiedzy ma formę: podręczników, artykułów, szkoleń itd. Jednak w kontekście systemu informatycznego, który ma zarządzać tą informacją, jest to bardzo istotna determinanta. Oznacza to, że jest ona zapisana językiem naturalnym

bez zastosowania formalnego opisu części składowych informacji oraz zaleceń, jak ją interpretować w trakcie przetwarzania.

W SZW taka informacja ma zwykle postać tekstową i reprezentuje wiedzę innych ludzi zgodnie z rys. 3. Niekiedy ma formę grafik, obrazków, dźwięków, filmów, którymi na razie zarządza się w bardzo ograniczonym stopniu, ponieważ brakuje skutecznych algorytmów przetwarzania takich informacji. Jednak zwykle nawet takie informacje nietekstowe są opisane dodatkowo tzw. metadanymi, czyli wspomnianą strukturą, zazwyczaj w postaci tekstu, aby łatwiej było nimi zarządzać.

Można wyróżnić kilka typów struktur informacji reprezentującej wiedzę w SZW (struktur jawnych):

1. Ze względu na dynamikę:

a) **statyczna** – struktura, którą zwykle ma się na myśli, używając pojęcia struktury informacji. Można ją zdefiniować jako opis wyższego rzędu specyfikujący sposób uporządkowania, grupowania, przetwarzania i interpretacji informacji. W przypadku informacji słabo ustrukturalizowanej może to być np.: szablon listu składający się z tytułu, nagłówka, treści i podpisu;

b) **dynamiczna** – jest opisem, w jaki sposób dana informacja ma przepływać przez system informatyczny i organizację, czyli jest związana z funkcją czasu. Może ona być opisana w systemie zarządzania przepływem pracy (*workflow*), np.: memo stworzone przez określonych pracowników jest recenzowane przez innych i po zatwierdzeniu rozsyłane do reszty firmy. Struktura dynamiczna informuje o historii informacji i jej aktualnym stanie.

2. Ze względu na źródło pochodzenia:

a) **człowiek** – wartości wprowadzane są przez użytkownika, który wprowadza lub tylko opisuje daną porcję informacji. W zależności od typu systemu rola ta jest różnie nazywana, np.: redaktor w CMS, ekspert w SZW, agent w *workflow*, inżynier wiedzy w systemie ekspertowym lub w bazie przypadków. W praktyce wprowadzenie takich wartości sprowadza się zwykle do wypełnienia odpowiedniego formularza (z wyjątkiem systemu ekspertowego).

b) **komputer** – wartości są wpisywane przez automatyczny algorytm. Jako przykład można tu podać kategoryzację, grupowanie (*clustering*), automatyczne tworzenie abstraktu, indeksowanie słowami kluczowymi, cięcie tekstu na fragmenty (*text tiling*), ekstrakcję wiedzy z informacji. Metod automatycznego wypełniania struktury używa się przede wszystkim, gdy ma się do czynienia z dużym zasobem informacji, którego nie można opisać ręcznie.

Dla informacji reprezentujących wiedzę w systemie zarządzania wiedzą najbardziej powszechnym standardem zapisu struktury K^{*} jest XML. Zatem wystarczy strukturę porównywaną komunikatów zapisać za pomocą jednego typu XML i wyliczyć minimalną liczbę węzłów w obu przypadkach.

Najważniejszą rolą struktury w aspekcie informacji reprezentującej wiedzę jest funkcja informacyjna. Dzięki niej:

1. **Łatwiej odnaleźć informację w bazie wiedzy** – jednym z celów wyszukiwania informacji jest interpretowanie struktury symboli składających się na

informację, czyli struktury informacji [Meadow, Boyce, Kraft 2000]. Warto pamiętać, że metody wyszukiwania starają się nie tylko analizować strukturę informacji zapisaną w sposób jawny w formie atrybutów, co jest głównym przedmiotem rozważań w tej publikacji, ale również strukturę wewnętrzną, wynikającą samistnie z treści informacji. W [Fervruci 2004] twierdzi się, że głównym rezultatem analizy informacji nieustrukturalizowanej jest przypisanie jej metadanych zawierających informacje o stosowanym języku, częściach informacji, składni, odniesień do słowników, powiązań z kategoriami itd. Ta uwaga odnosi się do wszystkich poniższych metod automatycznego przetwarzania informacji tekstowej.

2. Łatwiej utrzymać jakość informacji w czasie – np. w SZW w pewnej firmie porcje informacji są regularnie weryfikowane i duże znaczenie ma tutaj struktura dynamiczna opisująca zależności między informacjami i ich cykl życia³ [Szumiło, Żmigrodzki, Kilian 2004]. Każda odpowiedź musi być zweryfikowana, gdy spełniony jest jeden z poniższych warunków:

- minęło 60 dni od ostatniej weryfikacji,
- użytkownik, który zadał pytanie, ocenił odpowiedź na mniej niż 3,
- użytkownik o wyższych uprawnieniach, tzw. recenzent, zażądał powtórnej weryfikacji odpowiedzi.

3. Łatwiej zinterpretować i przyswoić informację przez odbiorcę – np. L. Drelichowski [2004] postuluje standaryzację porcji wiedzy poprzez wprowadzenie struktury modułu wiedzy i jego zawartości, co jest warunkiem koniecznym do efektywnego tworzenia, przyswajania i zastosowania wiedzy. Taka struktura definiuje długość dokumentu w stronach (6 stron, 10-12 stron, 18-20 stron) oraz podstawowe części tekstu:

- 1) „wstęp i cel opracowania.
- 2) podstawowe pojęcia. (...)
- 3) analiza dotychczasowych tendencji i przyszłych kierunków rozwoju. (...)
- 4) poszerzenie oceny określonych kierunków rozwoju. (...)
- 5) omówienie kluczowych czynników rozwoju. (...)
- 6) sprecyzowanie wniosków poznawczych. (...)
- 7) literatura” [Drelichowski 2004, s. 313].

Dzięki korzystaniu z powszechnej, jednoznacznie rozumianej terminologii możliwa jest poprawa komunikacji międzyludzkiej, wzrost jakości lokalizowania informacji, w konsekwencji stopień wykorzystania informacji i w końcu przyswajania wiedzy. Jest to postulat wprowadzania silniejszej struktury wewnętrznej. Inaczej ujmując, w kontekście zasobów informacyjnych głównym produktem transferu wiedzy jest porcja informacji reprezentująca wiedzę.

Na koniec należy zilustrować rozważania przykładami. W pewnym systemie zarządzania wiedzą [PYTHON...] stworzono notkę: „Zmiana uchwały zarządu. Dnia 2004-05-20 zarząd podjął uchwałę zawieszającą produkt XYZ bezterminowo.

³ Cykl życia porcji informacji to zbiór stanów, przez które może ona przechodzić, np.: w przygotowaniu, zgłoszona do publikacji, zatwierdzona, opublikowana.

Produkt zostaje zawieszony od 2004-05-31. Już podpisane umowy są realizowane bez zmian.”. Ta notka opisana jest następującym zbiorem atrybutów A (strukturą): TYTUŁ⁴ = „Zawieszenie produktu XYZ”, ABSTRAKT = „Uchwała została zmieniona. Przeczytaj pełną treść, aby dowiedzieć się więcej”, TREŚĆ jest taka, jak podano wyżej, DATA UKAZANIA SIĘ = „2004-05-20”, DATA WYGASNIĘCIA = „2004-06-20”, ADRESAT WIADOMOŚCI = wszyscy użytkownicy, AUTOR = „Jan Kowalski”, KATEGORIE = „Produkt XYZ”, CZY WYŚLAĆ BIULETYNEM = tak, CZY POKAZAĆ NA STRONIE GŁÓWNEJ SYSTEMU = tak.

Taka struktura opisuje komunikat zarówno statycznie, jak i dynamicznie. Dzięki niej mogą z nią zapoznać się pracownicy, których ona interesuje. Jednocześnie adresaci wiedzą, że notka jest skierowana właśnie do nich. Notka jest przyporządkowana do odpowiedniej kategorii tematycznej. Dzięki strukturze SZW może również wspierać utrzymanie jej aktualności w czasie. Na podstawie wzoru [8] wielkość tej struktury wynosi: (10 atrybutów) – $\log_2(10) = 3,32$.

Na przykład przepis, o którym mówi ta notka ma następujące atrybuty (strukturę): NUMER, TYTUŁ, TREŚĆ, KOMENTARZ, PRZEPIS NADRZĘDNY, UCHWAŁA, DATA PUBLIKACJI, DATA WEJŚCIA W ŻYCIE, WERSJA, AUTOR, ZAŁĄCZNIKI. Zatem jego struktura ma wielkość – $\log_2(12) = 3,58$. Załącznikiem do tego przepisu może być dokument, np. plik Word. Taki dokument ma następujące atrybuty (strukturę): NAZWA, OPIS, PLIK, DATA DODANIA, DATA AKTUALIZACJI, DATA WYGASNIĘCIA, KATEGORIA, AUTOR. Wielkość jej wynosi – $\log_2(8) = 3$.

W tych przykładach mowa tylko o strukturze zewnętrznej. Zarówno notka, jak i przepis mogą mieć dodatkowo strukturę wewnętrzną, wynikającą z ich treści, jednak ze względu na trudności jej pomiaru została ona pominięta. Jak widać, wszystkie te komunikaty mają strukturę podobnej wielkości, jednak różniącą się doбором atrybutów.

5. Wnioski

Spojrzenie na zarządzanie wiedzą z perspektywy infologicznej pozwala sprowadzić różnorodne informacje do wspólnego mianownika i pokazuje, że zdecydowana większość informacji istotnej dla zarządzania wiedzą jest słabo ustrukturalizowana. Słaby poziom ustrukturalizowania powoduje, że utrudnione jest zarządzanie nią za pomocą automatycznych algorytmów. Jednocześnie występuje duża różnorodność form informacji w SZW. Te czynniki w połączeniu z dużą ilością informacji obecnych w przedsiębiorstwie stwarzają poważną barierę dla efektywnego zarządzania wiedzą.

⁴ Nazwy atrybutów dla przejrzystości pisane są wielkimi literami. Atrybuty zgodnie z tym, co powiedziano wcześniej, są węzłami w XML.

Jednocześnie istnienie silnych struktur informacji reprezentującej wiedzę (w literaturze pisze się niekiedy o strukturach wiedzy) pozwala na jej łatwe kopiowanie i zwiększa stopień powtórnego wykorzystania [Prebst, Raub, Romhardt 2002; Perechuda 2005]. Zatem strukturalizacja informacji jest pożądanym działaniem w ramach zarządzania wiedzą. Potwierdza to duże zainteresowanie, jakie ostatnio przejawiają dostawcy rozwiązań zarządzania informacją słabo ustrukturalizowaną.

Docenienie roli struktury informacji reprezentującej oraz dokonanie jej pomiaru może pomóc w opracowaniu metod jej strukturalizowania. Dzięki temu zaś może stać się możliwe, z jednej strony, lepsze zarządzanie wiedzą, a z drugiej – połączenie metod przetwarzania informacji silnie i słabo ustrukturalizowanej, które dotychczas należą do odrębnych dziedzin.

Literatura

- Anderson J.R., *Uczenie się i pamięć. Integracja zagadnień*, Wydawnictwa Szkolne i Pedagogiczne, Warszawa 1998.
- Ark Group – witryna firmy, <http://www.ark-group.com>.
- Birch A., Malim T., *Psychologia rozwojowa w zarysie. Od niemowlęstwa do dorosłości*, Wydawnictwo Naukowe PWN, Warszawa 1998.
- Davenport T.H., Prusak L., *Working Knowledge*, Harvard Business School Press, Boston 1998.
- Drelichowski L., *Koncepcja metod edukacji i zastosowań technik informacyjnych w społeczeństwie wiedzy*, [w:] W. Abramowicz, A. Nowicki, M. Owoc, *Zarządzanie wiedzą w systemach informacyjnych*, AE, Wrocław 2004.
- Encyklopedia PWN 2003*, Wydawnictwo Naukowe PWN, Warszawa 2003.
- Ferruci D., Lally A., *Building an Example Application with the Unstructured Information Management Architecture*, „IBM Systems Journal” 2004 vol. 43, nr 3.
- Jones R.R., Bremdal B. A., Spaggiari C., Johansen F., Engels R., *Knowledge Management through Content Interpretation*, Halden 2000.
- Knowledge Management in the Internet Economy. Leveraging Knowledge for e-Business Success*, artykuł opublikowany przez Intel Corporation, 2001.
- Knowledge Management Magazine* – witryna czasopisma, <http://www.kmmagazine.com>.
- Kontzer T., *Content Overload*, Information Week Jan 19, 2004.
- Kugel R.D., *Unstructured Information Management, Intelligent BPM*, http://www.intelligentbpm.com//feature/2003/12/0312feat2_1.html, 2003.
- Langham M., *Mapping the Forest of Unstructured Information*, http://www.it-analysis.com/article_pf.php?articleid=3756, 2003.
- Meadow C.T., Boyce B.R., Kraft D.H., *Text Information Retrieval Systems*, Academic Press, San Diego 2000.
- Miller R., *Get Smart with Business Intelligence Software*, EContent, 2003/11, 2003.
- Nonaka I., Takeuchi H., *Kreowanie wiedzy w organizacji. Jak spółki japońskie dynamizują procesy innowacyjne*, Polska Fundacja Promocji Kadr, Warszawa 2000.
- Pai P., Miller L.L., Honavar V., Wong J., *Supporting Organizational Knowledge Management with Agents*, [w:] D. White, *Knowledge Mapping & Management*, IRM Press, London 2002.
- Perechuda K., *Dyfuzja wiedzy w przedsiębiorstwie sieciowym. Wizualizacja i kompozycja*, AE, Wrocław 2005.
- Probst G., Raub S., Romhardt K., *Zarządzanie wiedzą w organizacji*, Wydawnictwo Oficyna Ekonomiczna, Kraków 2002.

- PYTON Enterprise – witryna systemu, <http://www.pyton.pl>.
- Smith R.G., Farquhar A., *The Road Ahead for Knowledge Management An AI Perspective*, AI Magazine, Winter 2000, AAAI 2000.
- Stefanowicz B., *Informacja*, Szkoła Główna Handlowa w Warszawie – Oficyna Wydawnicza, Warszawa 2004.
- Stensmo M., Thorson M., *Unstructured Information Management. An Overview of the Enterprise Search*, Text Analysis and Visualisation Market, Infosphere, <http://www.infosphere.se>, 2003.
- Szumilo I., Żmigrodzki M., Kilian B., *Raport z pilotażu Systemu Zarządzania Wiedzą PYTHON dla DEF*, materiały z wdrożenia, Wrocław 2004.
- Tavris C., Wade C., *Psychologia. Podejścia oraz koncepcje*, Zysk i S-ka, Poznań 1995.
- Trombly M., *Content Management: Taming the Info Flood*, Securities Industry News, Volume XV, Number 31, August 11, 2003.

STRUCTURE OF INFORMATION REPRESENTING KNOWLEDGE

Summary

The article explains the role of information structure in knowledge management in the context of information science. It defines information structure as a set of attributes describing content of a piece of information. Then levels and strength of information structure are analyzed, and the process of building information structure is presented. At the end it is concluded that information representing knowledge in modern enterprises is mainly weak structured.

Jadwiga Sobieska-Karpińska – dr hab., prof. AE w Katedrze Komunikacji Gospodarczej Akademii Ekonomicznej we Wrocławiu.

Marcin Żmigrodzki – mgr, doktorant studiów zaocznych na Wydziale Zarządzania i Informatyki Akademii Ekonomicznej we Wrocławiu.