

Wiesława Gryncewicz

CHARAKTERYSTYKI I MIARY JAKOŚCI INFORMACJI

1. Wstęp

Jakość informacji trzeba mierzyć po to, aby wszyscy użytkownicy mieli pewność, że operują dokładnymi zasobami informacyjnymi. W przypadku niezadowolającej oceny należy tę jakość poprawiać i doskonalić. Koszty posługiwania się informacjami o niskiej jakości są bowiem wysokie (por. Gryncewicz 2003, s. 94-100). Jest to związane z koniecznością poprawiania lub ponownego wykonywania określonych operacji i pociąga za sobą stratę czasu ludzi, pieniędzy, materiałów i innych zasobów.

W niniejszym artykule zostaną opisane atrybuty i miary jakości informacji oraz autorska propozycja wskaźnika jakości informacji.

2. Definicja jakości informacji

L. English (1999, s. 22) wyróżnia inherentną i pragmatyczną jakość informacji. **Inherentną jakość informacji** utożsamia z dokładnością danych tworzących informację (*data accuracy*). Jest to stopień, w jakim dane te dokładnie odzwierciedlają, opisują rzeczywisty obiekt, którego dotyczą (*correctness of facts*). **Pragmatyczna jakość informacji** to stopień użyteczności informacji, stopień, w jakim informacja umożliwia odbiorcy osiągnięcie założonych celów skutecznie i efektywnie (*degree of usefulness*).

Posiłkując się powyższymi definicjami oraz innymi, zaczerpniętymi z licznych pozycji literaturowych dotyczących problematyki jakości oraz teorii informacji, w artykule przyjęto, iż **jakość informacji odzwierciedla stopień spełnienia przez nią wymagań użytkownika**. Stwierdzenie to łączy w sobie zarówno inherentne, jak i pragmatyczne cechy jakości informacji. Aby w pełni sprostać wymaganiom odbiorcy, określona informacja musi bowiem nie tylko dokładnie opisywać rzeczywistość, ale jednocześnie mieć inne cechy, które sprawia, że będzie użyteczna.

Ponadto powyższa definicja:

- stawia na centralnym miejscu użytkownika informacji,

- akcentuje to, że jakość informacji jest pojęciem subiektywnym uzależnionym od oczekiwań odbiorcy,
- wskazuje, że jakość informacji jest kategorią dynamiczną, zmienną w czasie, podobnie jak zmieniają się w czasie potrzeby użytkowników.

3. Charakterystyki jakości informacji

W literaturze przedmiotu nie ma jednolitego katalogu cech określających informację wysokiej jakości. Zarówno w publikacjach z lat 80. oraz 90., jak i w tych najnowszych, krajowych i zagranicznych, próbuje się ustalić listę atrybutów przynależnych temu pojęciu.

Niektórzy autorzy uważają, że na podstawie celów i funkcji zarządzania oraz zakresu działania odbiorców informacji można stwierdzić, iż informacja taka powinna być (por. Buśko i in. 1980):

- prawdziwa (wiarygodna), a więc obiektywnie zgodna z przedmiotem, który opisuje w danym czasie i miejscu,
- aktualna (nadażna), to znaczy, że powinna być dostarczona odbiorcy w czasie trwania sytuacji, której dotyczy, i opisywać ostatnie możliwe do zarejestrowania cechy stanów, zdarzeń i procesów,
- szczegółowa, a więc jej zakres powinien być na danym szczeblu zarządzania zgodny z klasą rozwiązywanych problemów decyzyjnych, przy czym należy pamiętać, że nadmiar informacji (jej zbytnia szczegółowość) jest równie szkodliwy jak niedobór,
- adresowalna, czyli skierowana do konkretnych odbiorców, którym jest obiektywnie niezbędna ze względu na zakres wykonywanej pracy,
- dokładna, to znaczy dostatecznie precyzyjna i odpowiednio ustrukturalizowana w granicach potrzeb odbiorcy i dopuszczalnego kosztu przetwarzania,
- niesprzeczna, a więc zgodna logicznie (semantycznie) oraz merytorycznie i formalnie z dziedziną przedmiotową, której dotyczy,
- dostępna, tj. możliwa do uzyskania w wymaganym czasie i miejscu, za pomocą jak najprostszych środków, akceptowanych przez użytkownika,
- zupełna (kompletna), czyli obejmująca swym zakresem całe zagadnienie, którego dotyczy dany problem decyzyjny,
- jednoznaczna, a więc identyfikująca i opisująca ściśle określony problem w sposób nie stwarzający problemów interpretacyjnych,
- tania (ekonomiczna), czyli uzyskana po jak najniższych kosztach.

Również z perspektywy zarządzania przedsiębiorstwem spogląda na jakość informacji J. Kisielnicki (1993, s. 33-37). Wymienia on takie atrybuty, jak:

- dyspozycyjność,
- aktualność,
- rzetelność,
- porównywalność,

- niezawodność,
- elastyczność,
- wydajność,
- czas reakcji,
- szczegółowość.

Takimi atrybutami powinny się charakteryzować informacje służące zaspokojeniu potrzeb informacyjnych menadżerów zarządzających przedsiębiorstwami.

Także M. Niedźwiedziński zauważył, że w literaturze specjalistycznej wymienia się, z różnym stopniem precyzji, kilkadziesiąt cech jakościowych informacji. Z tego bogatego katalogu wybrał i sam zwrócił uwagę tylko na takie atrybuty, jak (Niedźwiedziński 1987, s. 361-369):

- agregacja,
- aktualność,
- celowość,
- cenność,
- dokładność,
- dostępność,
- jednoznaczność,
- kompletność,
- opłacalność,
- wiarygodność,
- wierność,
- prawdziwość.

Dobór atrybutów jakości informacji zależy od przeznaczenia samej informacji oraz od jej potencjalnego użytkownika. U. Gupta (2000) uważa, że sukces przedsiębiorstwa w XXI w. zależy od sprawnego systemu informacyjnego. Jego zdaniem zbiory informacji występujące w takim systemie powinny charakteryzować się następującymi cechami:

- przydatnością informacji w sensie subiektywnym,
- związkiem z rozwiązywanym problemem (relewantność),
- terminowością,
- dokładnością,
- rozsądnym formatem (dostosowanym do użytkownika w kontekście rozwiązywanego problemu, nie wymagającym dodatkowego przetwarzania),
- kompletnością (a raczej wystarczalnością w opinii odbiorcy),
- dostępnością (z uwzględnieniem z jednej strony warunków bezpieczeństwa informacji, a z drugiej – zdolności percepcyjnych użytkownika).

W rzeczywistości jednak trudno o informację, która będzie miała wszystkie te cechy, poza tym, jak już wspomniano, jakość informacji jest nierozzerwalnie związana z jej przyszłym użytkownikiem i jego potrzebami oraz momentem, w którym będzie on ją wykorzystywał. Ciekawe podejście do tej problematyki

zapropował E. Kolbusz (1993, s. 143). Wprowadził on pojęcie **użyteczności**, jako syntetyczne określenie cech jakościowych informacji, mieszczące w sobie dowolną konfigurację wymienionych wcześniej atrybutów zależną od potrzeb odbiorcy. Informacja użyteczna to taka, która zawiera oczekiwany sens i ma cechę istotności. Oznacza to, iż w zbiorze wszystkich przypisywanych jej cech znajdują się również te, które są istotne z punktu widzenia celu, w jakim została dostarczona. Użyteczność danej informacji za każdym razem ocenia jej użytkownik.

Pewną próbę systematyzacji omawianych cech zaproponował również L. Floridi (2005). Stworzył on listę 27 atrybutów, które podzielił na 4 kategorie.

1. **Cechy modalne** (*modal properties*), odzwierciedlające stopień przystosowania informacji do wykorzystania. Do nich zaliczył w szczególności spójność informacji oraz jej faktyczne istnienie.

2. **Cechy humanistyczne** (*humanistic properties*), wśród których wyróżnił m.in. dokładność, integralność oraz bogactwo informacji.

3. **Cechy wyjaśniające** (*illuministic properties*), np. różnorodność form, dostępność informacji, możliwość jej przekazania różnym użytkownikom, jej systematyczną dostępność.

4. **Cechy konstruktywne** (*constructionist properties*), są to m.in. poprawność informacji, jej aktualność i normatywność.

Jeszcze inne podejście w zakresie systematyzacji zastosował L. English (1999, s. 141 i n.). Wyróżnił on inherentne i pragmatyczne cechy jakości informacji. **Cechy inherentne** (*inherent characteristics*) to cechy pierwotne, niezależne od sposobu użycia danych, opisujące cechy statyczne danej informacji. Są to m.in.:

- zgodność z definicją,
- kompletność,
- wiarygodność,
- zgodność z innymi źródłami,
- zgodność z rzeczywistością,
- dokładność (precyzyjność),
- stopień redundancji,
- paralelność.

Cechy pragmatyczne (*pragmatic characteristics*) natomiast określają stopień, w jakim informacje pozwalają użytkownikom efektywnie wykonywać ich zadania, a zatem dotyczą procesu dynamicznego. Należą do nich m.in.:

- dostępność,
- terminowość,
- zrozumiałość,
- zgodność informacji pochodnych z pierwotnymi,
- użyteczność,
- prawidłowość.

4. Zastosowanie podejścia infologicznego do opisu jakościowych cech informacji

Wiele z wymienionych w poprzednim punkcie artykułu atrybutów powtarza się u poszczególnych autorów lub jest inaczej nazywana. Może się również zdarzyć, że pod jednym pojęciem kryją się dwie nieco odmienne własności. Dlatego też, omawiając je, nie można polegać na ich intuicyjnym rozumieniu, lecz należy je krótko scharakteryzować. Opisy i interpretacje cech jakościowych są zazwyczaj mało precyzyjne i nieścisłe. Przyczyn takiego stanu rzeczy należy upatrywać w braku powszechnie akceptowalnej definicji informacji, a także w braku jednolitych zasad definiowania tych właściwości. Dlatego też część z nich odnosi się do informacji, część do metod i sposobów zbierania informacji, a część do sposobu interpretacji informacji przez użytkownika. Infologiczna interpretacja informacji ułatwia częściowe wyeliminowanie powyższych problemów. Jest to możliwe dzięki powiązaniu każdej pożądanej cechy jakościowej z określonymi elementami komunikatu, które występują w zapisie $K: = p(O, A, t, v)$ (Stefanowicz 2004, s. 17), gdzie:

K – komunikat będący nośnikiem informacji,

p – predykat opisujący obiekt O ,

O – obiekt będący przedmiotem opisu,

A – atrybut obiektu O ,

t – czas, w którym obiekt O jest rozpatrywany ze względu na wartość atrybutu A ,

v – wektor dodatkowych charakterystyk związanych z obiektem O , atrybutem A i czasem t .

Elementy te będą wykorzystywane w zaprezentowanych dalej definicjach atrybutów jakości informacji.

Sposób doboru cech jakościowych informacji oraz rozumienie treści każdej z nich zależą od użytkownika informacji. Jego cechy osobowościowe, wiedza oraz doświadczenie wpływają na postać katalogu atrybutów jakościowych i na ich wartościowanie. Poniżej przedstawiona zostanie analiza wybranych atrybutów jakości informacji przy zastosowaniu podejścia infologicznego:

- aktualności,
- dokładności,
- dostępności,
- jednoznaczności,
- kompletności,
- rzetelności,
- terminowości,
- użyteczności,
- wiarygodności,
- zrozumiałości.

Aktualność informacji. Zazwyczaj cechę tę interpretuje się jako zgodność informacji ze stanem rzeczywistym opisywanego obiektu, gdzie duże znaczenie ma czas (por. Pañkowska 2001, s. 15). To on bowiem sprawia, że obiekt O opisywany przez komunikat K ulega ciągłej transformacji, w wyniku której atrybut A zyskuje wciąż nowe (bardziej aktualne) wartości. Różnice te mogą być również wywołane zmianami w sposobie definiowania lub interpretacji samego atrybutu bądź całego obiektu. Jeżeli te różnice są nieistotne dla użytkownika, to taką informację możemy uznać za aktualną. W przeciwnym wypadku będzie to informacja nieaktualna.

Dokładność informacji. Dokładność można rozpatrywać dwojako:

- jako stopień uszczegółowienia informacji,
- w znaczeniu przyjmowanym w matematyce i odnoszonym do wartości mierzalnych.

W pierwszym znaczeniu odnosi się ona do tych elementów komunikatu K , które decydują o stopniu szczegółowości $I(K)$, to znaczy właściwego zdefiniowania obiektu O , atrybutu A oraz czasu t . W drugim ujęciu oznacza ona dokładność pomiaru wartości atrybutu A . Każda z tych danych może być podana z różnym stopniem dokładności, więc dokładność $I(K)$ jest wypadkową dokładności częściowych.

Dostępność informacji. Jest to stopień łatwości pozyskania przez zainteresowanego odbiorcę niezbędnej mu informacji $I(K)$.

Jednoznaczność informacji. Jest to cecha związana ze stosowaniem odpowiednio jednoznacznego języka i precyzyjnie zdefiniowanych pojęć. Jednoznaczność informacji w interpretacji infologicznej zależy od przyjętych zasad przedstawiania komunikatu K , czyli zasad oznaczania obiektów O i atrybutów A wraz z ich wartościami.

Kompletność informacji. W praktyce nie jest możliwe zgromadzenie wszystkich informacji na określony temat. Jest to spowodowane nie tylko ograniczeniami w postaci czasu i kosztów, ale także tym, że informacje są niewyczerpywalne. Dlatego to pojęcie jest interpretowane jako wystarczalność, czyli celem staje się pozyskanie tylu i takich informacji, jakie wystarczą do podjęcia racjonalnego działania. Należy zatem określić, jakie obiekty będą poddane badaniu, jakie ich atrybuty są niezbędne i w jakim przedziale czasowym będą analizowane. Tak uzyskaną informację uznaje się za kompletną.

Rzetelność informacji. W interpretacji infologicznej rzetelność informacji jest powiązana z rzetelnością procedur ich zbierania i przetwarzania. Zależy zatem od takich czynników, jak metoda zbierania informacji, metoda obserwacji obiektu O i poprawność jego identyfikacji, rodzaj źródła pozyskania wartości atrybutu A , właściwe sformułowanie pytań w ankietach i formularzach itp. Są to zatem elementy wektora v . Jeżeli w procesie zbierania informacji zostały dobrze dobrane metody i rzetelnie zrealizowane zaplanowane procedury, to nie ma podstaw, by kwestionować rzetelność informacji.

Terminowość. Jeżeli dana informacja $I(K)$ dociera w czasie wymaganym do zrealizowania określonego zadania i użytkownik zdąży ją wykorzystać do rozwiązania problemu Q , to znaczy, że była terminowa.

Użyteczność. Użyteczność informacji rośnie wraz ze zwiększaniem jej przydatności do rozwiązania problemu Q , z którym boryka się użytkownik. Zależy ona od sposobu agregowania i udostępniania informacji oraz jej aktualności.

Wiarygodność. Jest to stopień bliskości pozyskanej wartości atrybutu do jego wartości prawdziwej (faktycznej). Informacja wiarygodna to ta, w której wszystkie jej elementy składowe są prawdziwe. Prawdziwy jest obiekt, którego dotyczy, prawdziwe są wartości atrybutów i cały kontekst sytuacyjny.

Zrozumiałość informacji. Jest to cecha umożliwiająca zrozumienie informacji przez użytkownika bez szczególnych zabiegów dotyczących interpretacji semantycznej treści zawartych w komunikacie K . Język użyty w komunikacie musi być zgodny z tezaurem pojęciowym odbiorcy, a więc musi uwzględniać znane mu pojęcia i terminologię.

Wykorzystanie podejścia infologicznego do opisu jakościowych cech informacji nasuwa następujące wnioski:

- intuicyjna interpretacja cech jakościowych jest uściślona dzięki zastosowaniu definicji komunikatu K i jego elementów składowych,
- rozpatrywane cechy mają charakter względny, orzekanie o tym, czy dana informacja je ma czy nie, nie może dokonywać się bez odniesienia się do okoliczności, w jakich ta informacja będzie wykorzystana,
- uwzględnienie czynnika czasu (t) podkreśla dynamiczny charakter zarówno tych cech, jak i jakości określonej informacji, którą owe cechy opisują,
- tak zdefiniowane cechy ułatwiają obiektywizację oceny jakości konkretnych informacji.

5. Miary jakości informacji

Mając zdefiniowane wszystkie cechy jakości informacji, można przystąpić do ich pomiaru. **Poszczególne atrybuty jakości informacji w momencie ich oceny stają się bowiem miarami tejże jakości**, należy tylko zdefiniować skalę. B. Stefanowicz (2004, s. 108-109) proponuje, żeby zbiór pożądaných cech jakościowych oznaczyć jako C . Aby sprawdzić, czy konkretna informacja $I(K)$ ma jakąś cechę $c_i \in C$, należy sprawdzić, czy taką cechę ma odpowiedni element komunikatu K w stopniu satysfakcjonującym użytkownika. Następnie należy ustalić skalę S (od 1 do n), a wszystkie cechy C brane pod uwagę uporządkować stosownie do ich ważności. Wówczas dla każdej cechy $c_i \in C$ oblicza się dwa parametry:

1) v_i – miarę spełnienia przez informację wymagania sformułowanego przez użytkownika w postaci cechy c_i ;

$$v_i = (x - m_i)/x,$$

gdzie: x – ogólna liczba komunikatów,
 m_i – liczba komunikatów nie mających cechy c_i w stopniu satysfakcjonującym użytkownika;

2) w_i – wagę cechy c_i :

$$w_i = n - (k_i - 1)/n,$$

gdzie: n – maksymalna wartość w przyjętej skali oceniania,

k_i – grupa, do której została zaliczona cecha c_i .

Dysponując tymi wielkościami, można obliczyć łączny wskaźnik v jakości informacji:

$$v = (w_1 \cdot v_1 + \dots + w_n \cdot v_n) / (w_1 + \dots + w_n).$$

Sformułowany przez B. Stefanowicza wskaźnik pozwala łączyć pojęcie jakości z pojęciem informacji na wspólnej płaszczyźnie, jaką jest komunikat. Stwarza to podstawy do formułowania wniosków w zakresie jakości informacji oraz kierunków jej doskonalenia niezależnie od interpretacji poszczególnych cech jakościowych.

W artykule zaproponowano inną budowę wskaźnika jakości informacji. Jest ona oparta na propozycjach cytowanego autora. Otóż każdy komunikat będący nośnikiem istotnych informacji powinien zostać oceniony pod kątem wybranych przez użytkownika cech jakościowych w następujący sposób:

$$q_i = (o_1 + o_2 + \dots + o_n)/n,$$

gdzie: q_i – wskaźnik jakości komunikatu i ,

i – liczba komunikatów, przybiera wartości: 1, 2, ..., m ,

o_i – miara stopnia spełnienia przez komunikat i wymagania sformułowanego przez użytkownika w postaci cechy c_j ,

j – liczba cech będących przedmiotem oceny, przybiera wartości: 1, 2, ..., n .

Jeżeli w badanym procesie korzysta się z kilku komunikatów, to łączny wskaźnik jakości informacji dla wszystkich komunikatów występujących w całym procesie ma postać następującą:

$$Q = (q_1 + q_2 + \dots + q_m)/m.$$

Wskaźnik ten jest średnią wskaźników jakości poszczególnych komunikatów.

Powyższe podejście ma tę zaletę, iż pozwoli ocenić jakość informacji zawartych w poszczególnych komunikatach. Po przeprowadzeniu badania będzie można powiedzieć, jakie noty otrzymał dany komunikat, a nie tylko stwierdzić, ile komunikatów otrzymało oceny niezadowolające. Możliwa zatem będzie bardziej wnikliwa analiza każdego z procesów i dotarcie do źródeł wywierających wpływ na jakość informacji.

6. Zakończenie

Przedstawiona definicja jakości informacji oraz charakterystyka jej atrybutów wskazują na to, iż są to kategorie dynamiczne, uzależnione od odbiorcy, jego potrzeb i oczekiwań. Dlatego też miary jakości informacji odwołują się do wymagań sformułowanych przez użytkowników. Koncepcja infologiczna, posługując się definicją komunikatu, stworzyła dogodne warunki do badania oraz wyjaśniania funkcji i cech informacji, a także do analizy potrzeb informacyjnych jej odbiorców. Pozwoliła ona również na ukierunkowanie rozważań na kwestię odpowiedniej strukturalizacji informacji oraz uwzględniania różnych czynników, od których zależy sposób interpretacji informacji.

Dzięki temu, że zaproponowany wskaźnik jakości informacji jest oparty na definicji komunikatu, będzie możliwe ocenianie poszczególnych komunikatów i na ich podstawie formułowanie wniosków dotyczących jakości informacji zawartych w tych komunikatach. W przypadku komunikatów, które uzyskują oceny niezadowolające, należy indywidualnie rozpatrywać noty za poszczególne cechy jakościowe, wskazywać obszary, w których w pierwszej kolejności trzeba podjąć działania naprawcze, oraz szukać przyczyn i źródeł problemów.

Literatura

- Buśko B., Filipek H., Śliwieński J., *Wiarygodność informacji ekonomicznej w systemach informacyjnych*, PWE, Warszawa 1980.
- English L., *Improving Data Warehouse and Business Information Quality*, John Wiley & Sons Inc., New York 1999.
- Floridi L., *Information Ethics, Its Nature and Scope*, „Computer & Society” 2005 34(5).
- Gryncewicz W., *Ocena jakości informacji jako instrument zarządzania współczesnym przedsiębiorstwem*, [w:] *Instrumenty zarządzania we współczesnym przedsiębiorstwie*, red. K. Zimniewicz, AE, Poznań 2003, s. 94-100.
- Gupta U., *Information Systems: Success in the 21st Century*, Prentice Hall Canada Inc. 2000.
- Kisielnicki J., *Informatyczna infrastruktura zarządzania*, Wydawnictwo Naukowe PWN, Warszawa 1993.
- Kolbusz E., *Analiza potrzeb informacyjnych przedsiębiorstwa. Podstawy metodologiczne*, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin 1993.
- Niedźwiedziński M., *Cechy informacji – próba systematyzacji*, [w:] *Jakość danych w systemach informacyjnych*, red. J. Oleński. Seria: Systemy Informatyczne nr 1, Wydawnictwo OBSR, Warszawa 1987, s. 360-370.
- Pańkowska M., *Zarządzanie zasobami informatycznymi*, Difin, Warszawa 2001.
- Stefanowicz B., *Informacja*, SGH, Warszawa 2004.

INFORMATION QUALITY CHARACTERISTICS AND MEASURES

Summary

Information is of high quality if it meets users' needs. Accurate, complete and timely information is required for optimal business performance. This article defines information quality, its characteristics and measures based on infology. The authoress presents also her own proposition of information quality indicator.