

Digitalizacja a cyfryzacja

Wpisany przez Marek Zieliński

wtorek, 17 grudnia 2013 00:00 - Poprawiony wtorek, 17 grudnia 2013 12:48



Digitalizacja. Ilustracja wykonana z wykorzystaniem pracy Junior Melo udostępnionej przez Wikimedia Commons [[CC-BY-SA-3.0](#)])

W języku polskim pojawiło się nowe słowo, 'cyfryzacja'. Zostało ono spopularyzowane przez stworzenie w 2011 roku nowego ministerstwa Administracji i Cyfryzacji, obejmującego szeroki zakres działania, od mniejszości religijnych do informatyzacji. Słowo to zwiększyło tylko pogmatwanie, gdyż niektórzy zaczęli używać go zamiast słowa digitalizacja, które także jest używane w różnych znaczeniach. Czym więc jest digitalizacja, formatowanie cyfrowe i cyfryzacja?

Formatowanie cyfrowe

To, co dociera do naszych zmysłów ma najczęściej charakter sygnału ciągłego, t.j. takiego, który może przyjmować dowolne wartości w swojej domenie. Światło - zarówno w jego natężeniu jak i kolorze jest sygnałem ciągłym, takim jest też dźwięk (charakteryzujący się też natężeniem i tonem). Natomiast nuty albo tekst zapisany na papierze jest ciągiem dyskretnym, tj. ciągiem znaków, z których każdy jest wybrany z tylko ograniczonego zestawu możliwości.

Konwersja sygnału ciągłego na dyskretny, albo konwersja analogowo-cyfrowa jest podstawą procesu digitalizacji. Odbywa się przez tzw. próbkowanie, tj. podział jakiegoś obszaru (domeny) na dyskretne elementy, i zapisanie wartości sygnału w każdym elemencie. Dla dźwięku próbkowanie polega na zapisie intensywności sygnału dźwiękowego w odcinkach czasu. Dla obrazu dwuwymiarowego, obszar dzieli się na prostokąty (najczęściej kwadraty) i zapisuje wartość sygnału - w tym wypadku koloru - dla danego kwadratu (pixla). Dyskretny jest zarówno

podział obszaru na elementy, jak i wartości, jakie po próbkowaniu może przyjąć zapis sygnału.

Jak się to odbywa w praktyce? Obraz dwuwymiarowy (płaski), np. zdjęcie, dokument, stronę tekstu, zamieniamy na zapis cyfrowy używając skanera; można też użyć aparatu fotograficznego. Sygnał dźwiękowy, przekształcony najpierw na elektryczny w mikrofonie, zamieniamy na zapis cyfrowy za pomocą urządzenia zwanego przetwornikiem analogowo-cyfrowym (AC). Przy konwersji filmu na błonie filmowej łączymy obie techniki - skan dla obrazu i przetwornik AC dla ścieżki dźwiękowej. Film zapisany na taśmie magnetycznej jest najpierw odtwarzany, a więc zamieniany z powrotem na sygnał elektryczny, który w podobny sposób jest zamieniany na cyfrowy z rozdzieleniem obrazu i dźwięku. W każdym wypadku przekształcamy jakąś informację zapisaną na 'klasycznym' nośniku na ciąg zer i jedynek, który z użyciem określonego kodowania zostaje zapisany na dysku lub w innej pamięci komputera.

Należy tu od razu odróżnić materiał który jest w chwili powstania zapisany w postaci cyfrowej - "born digital". Zdjęcie aparatem cyfrowym lub elektroniczny zapis koncertu to są zasoby które powstały jako cyfrowe, mimo, że w procesie produkcji mogły odbywać się procesy konwersji.

W domenie języka którym posługują się bibliotekarze i archiwiści, na proces konwersji analogowo-cyfrowej najczęściej używa się określenia "reformatting" albo "digital reformatting" - zamiana oryginału na surogat cyfrowy.

Metadane

Plik cyfrowy zawierający jakiś zasób jest początkowo zupełnie pozbawiony metadanych. To tak, jakby na dysku mieć setki tysięcy plików (komputer na którym piszę ten tekst ma ich prawie pół miliona) o nazwach "abdez" "rs23s" "94mpX" - nie mielibyśmy żadnych szans na znalezienie czegokolwiek. Proces zbierania metadanych (danych o danych, w tym wypadku danych o zasobie cyfrowym) można porównać do katalogowania książek, z tym, że pliki cyfrowe są o wiele bardziej niesforne i nie poddają się łatwo prostym zasadom katalogowania stworzonym przez bibliotekarzy. Metadane są ważne dla zasobów domowych (dawniej wnuki dostawały w spadku albumy fotografii dziadków), a krytyczne dla archiwów i bibliotek których zadaniem statutowym jest zachowanie i udostępnianie zasobów. O metadanych pisałem już [poprzednio](#), standardy metadanych są częścią cyklu w tym blogu.

Należy tu jeszcze wspomnieć o transkrypcji tekstu. Zeskanowany dokument jest czytelny dla człowieka, ale dla komputera to tylko gęsta masa kropek o różnych kolorach. Metadane są niezbędne do tego, żeby odszukać dany zasób, i często pomagają w tworzeniu indeksów

(tematy, osoby, miejsca, daty itp). Można jednak pójść dalej, i metadane uzupełnić o transkrypcję całego tekstu doprowadzając go do postaci zrozumiałej przez komputer. Można to zrobić ręcznie, ale zautomatyzowanie tego procesu jest możliwe dzięki technice optycznego rozpoznawania obrazu (OCR). Poddanie np. książki procesowi OCR (i usunięcie błędów, bo technika nie jest 100% dokładna) pozwala na jej dostarczenie w postaci cyfrowej, w wielu różnych formatach, w zależności od potrzeb, a także na przeszukiwanie zawartości całego tekstu.

Digitalizacja

Czym więc jest digitalizacja? Jest to cały proces przeróbki zasobu analogowego na cyfrowy, składający się z przygotowania, formatowania, opisu (zbierania metadanych) i udostępnienia. Proces digitalizacji może wyglądać nieco inaczej w zależności od rodzaju zasobu, tu opiszę ten proces dla archiwów, parafrazując fragment autorytatywnego dokumentu NARA (Narodowej Administracji Archiwów i Zasobów USA):

Digitalizacja to seria czynności których końcowym efektem jest kopia cyfrowa dostępna dla użytkowników poprzez Internet lub innymi kanałami, przez długi okres czasu. Na czynności te składa się:

- Identyfikacja i selekcja dokumentów. Jest to szczególnie ważne w archiwach, które często dostają pudła lub teczki zawierające wiele dokumentów, nie koniecznie w właściwej kolejności, często z pomieszzanymi stronami.
- Przygotowanie dokumentów, porządkowanie, konserwacja, paginowanie.
- Zbieranie podstawowych metadanych - zarówno technicznych jak i opisowych, wystarczających do znalezienia dokumentu i dostarczających podstawowych danych o kontekście, w którym ten dokument występuje.
- Konwersja cyfrowa.
- Kontrola jakości kopii cyfrowych oraz metadanych.
- Dostarczenie użytkownikowi (w postaci elektronicznej) dostępu do dokumentów.
- Utrzymanie kopii cyfrowych i metadanych, kopii zapasowych i planowanie na przyszłość (konwersje formatów i nowe urządzenia)

Tak jak biblioteka to nie magazyn książek (choć spełnia ona także funkcje magazynowe) tak digitalizacja to nie skanowanie (choć konwersja cyfrowa jest częścią procesu digitalizacji).

Cyfryzacja

Trudno znaleźć użyteczną definicję cyfryzacji - słowo to jest używane w wielu kontekstach w sposób dość dowolny. Przyjrzyjmy się więc, czym zajmuje się Ministerstwo Cyfryzacji. W programie "Polska Cyfrowa" ministerstwo definiuje trzy "osie": dostęp do szybkiego internetu, rozwój dostępnych w sieci e-usług i zasobów i kompetencje cyfrowych umiejętności obywateli. Pod hasłem "Kluczowe działania zintegrowanej informatyzacji" wymienia takie zadania jak zapewnienie ram organizacyjnych budowy systemu informacyjnego państwa, uzyskanie interoperacyjności rejestrów publicznych czy uruchomienie Państwowej Chmury Obliczeniowej. Widać więc wyraźnie, że pojęcie cyfryzacji jest tu rozumiane szeroko jako działania w celu zwiększenia dostępności Internetu i jego zasobów dla obywateli i wbudowanie mechanizmów elektronicznych w administrację państwową. Nie ma to więc nic wspólnego z digitalizacją.

Podsumowanie

Digitalizacja jest jednym z bardzo ważnych, a może najważniejszych zadań archiwów i bibliotek, wynikająca z powołania tych organizacji (zabezpieczenie i dostęp), w połączeniu z konserwacją i zabezpieczaniem oryginałów zasobów. Przyszłość, z czym zgadza się rząd powołując Ministerstwo Cyfryzacji, jest w dostępie elektronicznym do zasobów, w tym do książek i archiwów. Digitalizacja to cały proces zamiany istniejących zasobów analogowych w równoważne i dostępne zasoby cyfrowe, na co składa się m.in. obróbka oryginałów, konwersja cyfrowa, zbieranie metadanych i udostępnianie.

Czytaj więcej

- [Strategia Digitalizacji dla dostępu publicznego 2007-2016 \(NARA\)](#)
- [Wytyczne do tworzenia kolekcji cyfrowych, 2nd ed. Susan Schreibman \(pdf\)](#)
- [Formatowanie cyfrowe \(Wikipedia\)](#)
- [Metadane \(Wikipedia\)](#)

(wszystkie źródła w języku angielskim)

Marek Zieliński

Artykuł ukazuje się jednocześnie w *Blogu archiwistów i bibliotekarzy Instytutu Piłsudskiego*

Digitalizacja a cyfryzacja

Wpisany przez Marek Zieliński

wtorek, 17 grudnia 2013 00:00 - Poprawiony wtorek, 17 grudnia 2013 12:48

Może Cię też zainteresować

[Wstęp do standardów metadanych](#)