

Zegar astronomiczny w Pradze
By Steve Collis from Melbourne, Australia (Astronomical Clock Uploaded by russavia) [CC BY 2.0], via Wikimedia Commons

W jednym z poprzednich wpisów na blogu "Czy umiemy pisać daty?" omawiałem podstawy uniwersalnej notacji czasu i dat, zdefiniowanej w międzynarodowym standardzie ISO 8601 i jego uproszczonej wersji konsorcjum W3C. Od tego czasu Biblioteka Kongresu Amerykańskiego zakończyła prace nad rozszerzonym standardem, Extended Date/Time Format (EDTF) 1.0. Większa część EDTF dotyczy zapisu nieprecyzyjnych dat. Taka niedokładna lub nieprecyzyjna informacja dotycząca czasu występuje często w zapisach wydarzeń historycznych, np. w archiwach czy naukach bibliotecznych. Standard ISO 8601 nie pozwala na wyrażenie takich konceptów jak "w przybliżeniu rok 1962", "któryś rok pomiędzy 1920 a 1935" czy "wydarzenie miało prawdopodobnie miejsce w roku 1938, ale nie jesteśmy tego pewni". Standard EDTF pozwala na zapisanie w postaci zrozumiałej przez komputer takich konceptów, wypełniając potrzeby istniejące w wielu polach wiedzy mających do czynienia z metadanymi o charakterze historycznym.

Mimo tego, że standard EDTF jest stosunkowo nowy i nie ma zbyt wiele narzędzi programowych pomagających wprowadzać takie dane, sądzę, że warto jest zaznajomić się z tą

nowa notacją i używać jej w miarę możliwości

Definicje

Chciałbym rozpocząć dyskusję kilkoma definicjami; symbole pojawiające się przy definicjach będą opisane dalej.

Precyzja

Precyzja jest miarą zakresu, wewnątrz którego mieści się ‘prawdziwa’ wartość [1]. Precyzja jest jednoznacznie zdefiniowana w wyrażeniach daty i daty/czasu. Jeśli wydarzenie miało miejsce w roku 1318, zapis taki posiada precyzję jednego roku (mogło mieć miejsce w dowolnym czasie w ciągu tego roku). Jeśli podamy 1318-05, zwiększamy precyzję do jednego miesiąca, a 1945-09-15 posiada precyzję jednego dnia, itp [2]. W EDTF możemy rozszerzyć tę definicję określając precyzję dziesięcio- lub stulecia używając symbolu x (patrz 'precyzja maskowana' poniżej).

Przybliżona (~)

Szacunek który jest prawdopodobnie prawdziwy lub zbliżony do prawdziwego, gdzie stopień ‘zbliżenia’ zależy od danej aplikacji.

Niepewna (?)

Nie jesteśmy pewni wartości zmiennej (w naszym przypadku daty lub czasu). Niepewność jest niezależna od precyzji. Źródło informacji może samo nie być wiarygodne, możemy mieć do czynienia z kilkoma różnymi szacunkami i nie dość informacji aby zdecydować między nimi, itp.

Nieokreślona (u)

Wartość nie jest podana. Punkt w czasie może być nieokreślony ponieważ jeszcze nie nastąpił, bo jest tajny, nieznany lub z innych powodów.

Funkcjonalność EDTF

Standard EDTF rozszerza ISO 8601 w trzech poziomach. Poziom 0 jest zgodny z ISO 8501 i jego ograniczeniami w W3C. Poziom 0 jest opisany szczegółowo w poprzednim wpisie blogu i nie będę go tu powtarzał. Poziom 1 i 2 rozszerza ten standard wprowadzając dodatkowe

możliwości oznaczenia precyzji, przybliżenia, niepewności i nieokreślenia w datach. Jakkolwiek poziom 2 daje standardowi większą moc wyrazu, nie widzę powodu aby programiści nie wdrożyli obu poziomów w aplikacjach.

Pora roku

Zamiast miesiąca w ciągu rok-miesiąc można użyć liczb odpowiadających porom roku: 21 (wiosna), 22 (lato), 23 (jesień) i 24 (zima).

- **2014-21** (wiosna 2014)

W poziomie 2 można dodatkowo uściślić porę roku używając symbolu ^ jak w przykładzie: 2014-21^southernHemisphere. Jednakże słownik uściśleń nie został wyspecyfikowany w standardzie.

Data przybliżona i niepewna

Symbol **?** jest używany do wskazania na niepewną datę, a symbol ~ na przybliżoną. Oba symbole mogą być użyte osobno lub razem (w tym ostatnim przypadku znaczy to, że data jest przybliżona a nawet i to jest niepewne)

W poziomie 1 symbole **?** i ~ mogą być umieszczone tylko na końcu daty, i stosują się do całej daty:

- **1945?** (rok niepewny)
- **1945-03-12**~ (data przybliżona)
- **1945-03?**~ (rok-miesiąc przybliżony i niepewny)

W poziomie 2 dowolna część daty (ale tylko cały rok, cały miesiąc lub dzień) może być zaznaczona jako niepewna lub przybliżona. Symbol ma zastosowanie do części zapisu po lewej stronie od symbolu, i można użyć nawiasów aby wydzielić jakąś część zapisu.

- **1816?-05-25** (dzień i miesiąc znany, rok niepewny)
- **1816-05~-25** (dzień znany, rok i miesiąc przybliżone)

- **1816-(05)?-25** (tylko miesiąc niepewny, rok i dzień znany)
- **1816-(05-25)?** (rok znany, miesiąc i dzień niepewny)
- **1816?-05-25**~ (miesiąc znany, rok niepewny, dzień przybliżony)
- **(1816-(06)**~**)?** lub **1816?-(06)?**~ (rok niepewny, miesiąc przybliżony i niepewny)
- **1816-22**~ (pora roku przybliżona “około lata 1816”)

Data nieokreślona

Litera **u** może być użyta zamiast cyfry w dacie, jeśli ta część daty jest nieokreślona.

W poziomie 1 tylko cyfry po prawej stronie daty mogą być zastąpione. Drugim ograniczeniem jest to, że tylko jedna lub 2 cyfry w roku, i dokładnie dwie cyfry w miesiącu lub dniu mogą być zastąpione.

- **191u** (nieokreślony rok w latach 1910-tych)
- **19uu** (jakiś rok w latach 1900-ych)
- **1915-uu** (jakiś miesiąc w roku 1915)
- **1915-03-uu** (jakiś dzień w marcu 1915)
- **1915-uu-uu** (jakiś dzień w roku 1915)

W poziomie 2 te ograniczenia są zniesione, i **u** może zastąpić dowolną cyfrę w dacie. Jak zawsze, używamy liczby segmentów (rok, rok-miesiąc, rok-miesiąc-dzień) aby określić precyzję.

- **13uu-01-15** (15 stycznia w 1300-ych)
- **13uu-01-uu** (jakiś dzień w styczniu któregoś roku w 1300-ych)
- **13uu-01** (styczeń któregoś roku w 1300-ych)

Rozszerzony okres czasu

W okresie czasu (składającego się z dwóch dat przedzielonych znakiem /) możemy użyć słowa **unknown** zamiast daty początkowej lub końcowej, aby wskazać, że ta część jest nieznana. Możemy użyć słowa **open** zamiast końcowej daty aby wskazać, że okres czasu jest otwarty. Poza tym, symbole wskazujące na datę przybliżoną, niepewną i nieokreśloną (ten ostatni tylko w poziomie 2) mogą

też być użyte.

W poziomie 1 można użyć unknown i open; symbole ~ i ? zaś tylko na końcu każdej z dat.

- **1945-02-11/unknown** (okres rozpoczęty 2 lutego 1945, koniec nieznany)
- **2015-01-20/open** (okres rozpoczęty 20 stycznia 2015, bez daty końcowej)
- **1825~/1918-05** (okres rozpoczęty w przybliżeniu w 1825 i zakończony w maju 1918)

W poziomie 2 można użyć symboli ~, **?** i **u** w dowolnym miejscu w obu datach.

- **2012-(06)?-01/2015-06-nn** (okres rozpoczęty 1 czerwca 2016, gdzie miesiąc jest niepewny i zakończony w nieokreślonym dniu w czerwcu 2015)

Rok przekraczający 4 cyfry

Jeśli zapis roku jest dłuższy niż 4 cyfry (i tylko wtedy), poprzedzamy rok literą **y** (nie używamy w takim wypadku miesięcy ani dni). W poziomie 2 można dodatkowo użyć formy wykładniczej dla roku, oraz opcjonalnie dodać precyzję w postaci litery
p
po której podaje się liczbę cyfr znaczących.

Poziom 1

- **y-1700015** (rok minus 1700015)

Poziom 2

- **y17e5** (rok 1700000)

W poziomie 2 występują trzy dodatkowe, bardzo przydatne funkcje: precyzja maskowana i dwie listy dat.

Precyzja maskowana

W poziomie 0 wyrażamy precyzję zapisu daty przez określenie roku-miesiąca-dnia (precyzja 1 dnia), roku-miesiąca (precyzja miesiąca) i roku (precyzja roku). W poziomie 2 można zastąpić literą **x** ostatnią albo dwie ostatnie cyfry roku aby wskazać jeszcze mniejszą precyzję, dziesięciolecia i stulecia.

- **198x** (lata 1980-te)
- **19xx** (lata 1900-ne)

Należy zwrócić uwagę na to, że określenie ‘lata 1900-ne’ jest zbliżone ale nie identyczne z 20 wiekiem; 20 wiek zaczyna się w roku 1901 i kończy w 2000, podczas gdy lata 1900-ne zaczynają się w roku 1900 i kończą w 1999.

Jedna ze zbioru

Nawiasy kwadratowe **[** i **]** otaczają listę dat z której wybieramy tylko jedną. Wewnątrz nawiasów daty oddzielone są przecinkami, albo dwukropkiem oznaczającym wszystkie dat pomiędzy datami które oddziela dwukropek, włącznie z datami końcowymi. Różne elementy listy mogą mieć różną precyzję

- **[1821,1822,1830..1832]** (jeden rok z: 1821, 1822, 1830, 1831, 1832)
- **[..1935-11-15]** (15 listopad 1935 roku lub jakaś wcześniejsza data)
- **[1510-12..]** (grudzień 1510 lub jakiś późniejszy miesiąc)
- **[1725,1726-12]** (rok 1725 albo grudzień 1925)

Daty wielokrotne

Nawiasy klamrowe **{** i **}** otaczają listę, której wszystkie elementy powinny być uwzględnione. Dla kolejnych dat, oznacza to wszystkie daty w zbiorze, nie zakres dat. Przykładowo, możemy opisywać jakieś zdarzenie które powtarzało się w kilku różnych datach.

- **{1970-12..1972-12,1973-11}** (wydarzenie powtarzało się w grudniu 1970, w grudniu 1971, w grudniu 1972 i w listopadzie 1973)
- **{1950,1951-05}** (rok 1950 i maj roku 1951)

Jak używać EDTF

Nie ma jeszcze wielu implementacji czy przykładów użycia EDTF. Na stronie www Biblioteki Kongresu można znaleźć przewodnik po przydatnym oprogramowaniu używającym EDTF. Serwis weryfikacji zapisów pozwala na testowanie swoich rozwiązań przy nauce EDTF.

Wpisywanie daty sformatowanej według zasad EDTF w formularzach webowych nie zawsze jest możliwe, jeśli formularz egzekwuje standardowa datę lub podsuwa kalendarz. EDTF jest jednak o wiele lepsze niż tekstowy opis niepewnych lub rozszerzonych dat, gdyż zapis tekstowy rzadko może być zrozumiany przez komputery. Być może warto prosić programistów o takie pola w formularzach, korzystając z przykładów instytucji które już wdrożyły EDTF [4]. Jako plan minimum, EDTF powinien być używany w polach tekstowych zamiast słownego opisu.

Marek Zieliński

Artykuł ukazał się 1 lipca 2015 w *Blogu archiwistów i bibliotekarzy Instytutu Piłsudskiego*

Przypisy

[1] Dokładność i precyzja metod pomiaru - artykuł w Wikipedii
[2] Precyzja różni się od dokładności, która określa, jak pewni jesteśmy, że dane zdarzenie miało miejsce np. w danym roku.

Czytaj więcej

- Opis standardu EDTF na stronie Biblioteki Kongresu
- Używanie Rozszerzonego Formatu Daty i Czasu (EDTF) w ‘mark.e.philips journal’
- Lekcje z wdrażania Rozszerzonego Formatu Daty i Czasu w dużej bibliotece cyfrowej

- Przewodnik po oprogramowaniu EDTF

Może Cię też zainteresować

- Czy umiemy pisać daty?

- Co jest na odwrocie zdjęcia cyfrowego?

- Standardy metadanych: Dublin Core

- Czy jesteś GLAM?