

Zawsze od rana do środy
– Semantyka i normalizacja quasi-okresowych
określeń czasu w języku polskim

Jakub Kozakoszczak

Heinrich Heine Universität Düsseldorf, Uniwersytet Warszawski

5 listopada 2018

Plan

- 1 Motywacja
 - Projekt: system dialogowy
 - Założenia języka
- 2 Badane wyrażenia
 - Kontekst „Mam czas zawsze _”
 - *Piątek a piąty dzień tygodnia*
 - Kontekst „Mam czas zawsze od _”
- 3 Reprezentacja wiedzy czasowej
 - Przegląd istniejących rozwiązań
 - Relacje i operacje na warstwach
- 4 Propozycja
 - Warstwy podwójne
 - Operacje na warstwach podwójnych

- Projekt badawczo-rozwojowy współfinansowany z Europejskiego Funduszu Rozwoju Regionalnego, Działanie 1.2 „Działalność badawczo-rozwojowa przedsiębiorstw” nad wykorzystaniem języka naturalnego w innowacyjnej usłudze rezerwacji online.
- Projekt zmierza do zbudowania systemu komputerowego, który jest w stanie komunikować się jak życzliwy i pomocny człowiek.

- Implementowana reprezentacja znaczenia wypowiedzi osób rezerwujących usługi.
- Chatbot działa na podstawie ręcznie pisanych reguł.
- Reprezentacja maksymalnie czytelna dla człowieka.
- Reprezentacja blisko języka ze względu na poprawki i dopowiedzenia. Niedościęgnięty wzór: Leban et al. 1986.

- Część języka, która dostatecznie często realizowana jest w rezerwacjach.
- Dostosowanie do inkrementacyjnego sposobu mówienia:
koło 17 w piątek w przyszłym tygodniu
- Ograniczona warstwa inferencyjna. Reprezentowanie dostępności przez normalizację do przedziałów kalendarzowych i wnioskowanie przez proste przecinanie z innymi przedziałami.
- Podejście teorio-modelowe. Przykład ekspresywnego podejścia symbolicznego: (Niézette, Stevenne 1992).

Wyrażenia przyjmowane przez kontekst „Mam czas zawsze _”:
15 po, o 18, wieczorem, w środę, w weekend, trzeciego, latem, w przerwie meczu
Te same znaczenia można uzyskać pomijając „zawsze”.

	dzień roboczy	dzień upadku Muru	piątek	dzień tygodnia
\forall	każdy dzień roboczy	*każdy dzień upadku Muru	każdy piątek	każdy dzień tygodnia
kiedy?	w dzień roboczy	w dzień upadku Muru	w piątek	*w dzień tygodnia
zawsze	zawsze w dzień roboczy	*zawsze w dzień upadku Muru	zawsze w piątek	*w dzień tygodnia

„Mam czas _”:

15 po, o godz. 18,

Do czego dokładnie odnosi wyrażenie wskazujące godzinę?

- Do pełnej specyfikacji godziny i minuty (przeważnie).
- Do całej godziny zegarowej:

Codziennie o tej samej godzinie i minucie.

Reprezentujemy i normalizujemy *osiemnastą trzydzieści* jak przecięcie godziny zegarowej z krótszym przedziałem minutowym.

„Mam czas (**teraz**) _”:

15 po, 18, wieczorem, w środę, w weekend, trzeciego, latem, w przerwie meczu

- Skoro *każdy piątek* to po prostu *piąty dzień każdego tygodnia*, to *ten piątek* to *pierwszy piątek nieprzeszłości*.
- Znaczenie **zawsze** jako podstawa do reprezentacji obu znaczeń.
- Ale czy *piątek* to *piąty dzień tygodnia*?

- *Najbliższy piątek* ma tę samą ekstensję we wszystkich obszarach kulturowych, a *najbliższy piąty dzień tygodnia* różne.
- *Pierwsza* to druga godzina dnia. Jeden dzień w roku *drugiej* godziny.
- W 11. dniu listopada.
 - *W jedenastym listopada.
 - *W jedenastym.
- Kolejność: dzień drugi listopada, godzina pierwsza. Por. „Gol padł w minucie trzynastej meczu.”
- W dniu trzeciego listopada.
- *Wieczór* to które co czego?
 - Zawsze od 18 do wieczora.
 - *Zawsze od rana do środy.

„Mam czas zawsze od (do) _”:

*15 po, 18, wieczora, środy, weekendu, trzeciego, lata, przerwy
meczu*

- **Zawsze od dnia roboczego.*
- *Od środy* kończy się w niedzielę?
- *Od i do* reprezentowane i normalizowane jako niezależne operatory dające pożądaną przedział od-do w swoim przecięciu.
 - Odporność na dopowiedzenia.
 - To działa tylko wewnątrz cyklu.

Reprezentacja znaczeń wyrażen quasi-okresowych – odnoszących do okresowo powtarzających się przedziałów z możliwymi regularnymi zaburzeniami:

Collection (Leban et al. 1986): Lista wypukłych przedziałów (np. dni) lub innych kolekcji. Np. dni miesiąca to lista kolekcji, z których każda obejmuje dni jednego miesiąca).

- Dicing, slicing

Time sequence (Segev, Shoshani 1987): ciąg \langle atrybut obiektu, (dane czasowe, wartość atrybutu) \rangle^* uporządkowany czasowo.

Slice (podstawowy w Níezette, Stevenne 1992): Zbiór rozłącznych przedziałów pokrywających cały czas (*calendar*) lub wynik operacji *slicing* na innym *slice* lub suma teoriomnogościowa paru *slices*.

- Wykrawanie, wykrój

Granularity (Bettini et al. 1998), również *layer (of granularity)* (Euzenat, Montanari 2005): Funkcja mapująca liczby całkowite na zbiory punktów czasowych zachowująca ścisły porządek czasowy. Na początku i końcu mogą być spójne podciągi zbiorów pustych.

- Słownik środowiskowy, następca „The Consensus Glossary of Temporal Database Concepts”
- Frakcja, warstwa.

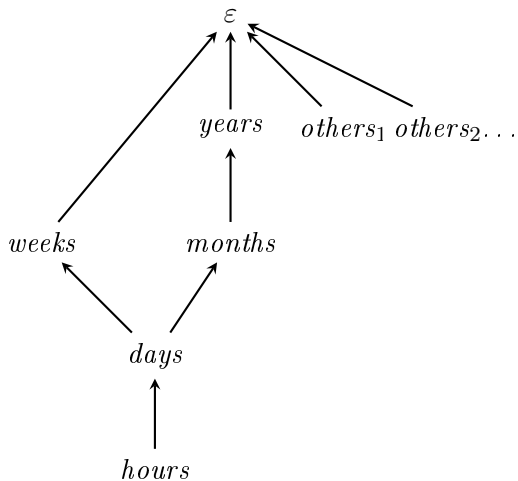
- **Slicing** (oryginalnie dicing + slicing) Czteroargumentowa operacja wykrawania.
- **Finer Than** Warstwa jest drobniejsza od drugiej jeśli każda jej cząstka zawiera się w jakiejś cząstce drugiej
- **Image** Obraz warstwy to suma mereologiczna jej cząstek

- Podstawowy obiekt: Warstwa podwójna składająca się z warstwy drobnej i warstwy grubej.

$$layer = \langle layer_{\text{fine}}, layer_{\text{coarse}} \rangle$$

- Drobna warstwa decyduje modeluje ekstensji wyrażenia, gruba decyduje o jego zachowaniu w kontekstach typu „zawsze od”.
- Dla wyrażen nie przyjmowanych przez kontekst „zawsze od” warstwa gruba
 - jest identyczna z drobną, np. *Days*. Inna warstwa podwójna uzyskana z tej przez wykrojenie lub przecięcie może być operandem „zawsze od”.
 - przyjmuje wartość ε blokującą kontekst „zawsze od”.
- Wykrawanie wyliczane jest na drobnych warstwach i zwraca parę \langle wyliczona warstwa, kres górny grubych warstw \rangle według porządku warstw grubych.

Porządek warstw grubych



Fridays =

$$\langle 5/Days_{\text{fine}} : \text{during} : Weeks_{\text{fine}}, \sup(\{Days_{\text{coarse}}, Weeks_{\text{coarse}}\}) \rangle = \\ \langle 5/Days_{\text{fine}} : \text{during} : Weeks_{\text{fine}}, Weeks_{\text{coarse}} \rangle$$

- **Begin** reprezentuje „zawsze od”.
- **End** reprezentuje „zawsze do”.
- **After** reprezentuje „zawsze po”.
- **Before** reprezentuje „zawsze przed”.
- Funkcja `coarse_grain` mapuje drobną cząstkę na grubą.
- Operacje są zdefiniowane, jeśli warstwa gruba nie jest pusta oraz funkcja `coarse_grain` jest zdefiniowana.

$$\text{coarse_grain}(g_f, l_c, l_f) = \\ \neg g_c (g_c \in l_c \wedge g_f = \text{Image}(l_f) \sqcap g_c)$$

- Funkcja `coarse_grain` jest niezdefiniowana (również) gdy $l_c = \varepsilon$

$$\begin{aligned} \text{begin}(\textit{layer}) = \\ \langle \{ \langle g_f^\alpha, \text{coarse_grain}(g_f, \textit{layer}_{\textit{fine}}, \textit{layer}_{\textit{coarse}})^\beta : g_f \in \textit{layer}_{\textit{fine}} \rangle \}, \\ \textit{layer}_{\textit{coarse}} \rangle \end{aligned}$$