

## Próba definicji systemu wstępnie niesformalizowanego

### Wstęp

Celem tej notatki jest próba określenia problemów, które można postawić w związku z analizą za pomocą maszyn matematycznych dowolnych tekstów.

Nie należy się łudzić, że przy tak ogólnie postawionym problemie da się uzyskać jakieś praktycznie wyniki. Uważam, że punktem wyjścia do tak postawionego zadania powinny być konkretne przykłady zastosowań, które być może w przyszłości dałoby się ująć w jakiś jednolity system. Tym niemniej sądzę, że wstępna ogólna analiza tego problemu może oddać pewne usługi pod warunkiem, że dalsze prace w tym kierunku poprowadzi się pod kątem konkretnych zastosowań.

### 1. Klasyfikacja tekstów

Uważam, że w skład badań interesującego nas kierunku powinny wchodzić jakieś próby chociażby podzielenia tekstów, które mogą być przedmiotem maszynowej obróbki - na różne kategorie. Orientacyjnie podam kilka znanych mi przykładów tego rodzaju tekstów.

#### a/ Teksty niezidentyfikowane

Przykładem tego rodzaju tekstów mogą być zaszyfrowane wiadomości lub napisy w nieznanym języku. O tekstach tego rodzaju nieamy /czy zakładamy/, że są one wyrażeniami jakiegoś języka /a więc posiadają prawidłową budowę/ oraz często posiadamy częściowe informacje o znaczeniu tekstu /lub prawdopodobnym znaczeniu/.

#### b/ Teorie

W uproszczeniu teksty tego rodzaju są zbiorem /albo raczej ciągłem/ zdań oznajmujących dotyczących określonego tematu. Np. opis geograficzny terenu, opis własności ciała fizycznego, pojęcia matematycznego są to przykłady teorii. Wśród teorii wyróżniamy teorie dedukcyjne, w których na postać tekstu narzucone są pewne ograniczenia.

Mianowicie teoria jest dedukcyjna jeżeli jej zdania są powiązane pewnymi zależnościami logicznymi. Przykładem mogą tu być oczywiście teksty matematyczne czy też wywody prawnicze - bądź też jakiegokolwiek inne wywody uzasadniające.

c/ Kodeksy

Przykładem tego rodzaju tekstu jest kodeks prawny /cywilny, karny/. W tekście takim należy wyróżnić conajmniej dwa rodzaje sformułowań: definicje oraz normy. Definicje określają tu pewne relacje między przedmiotami i podmiotami prawa, normy zaś określają pewne postępowania, gdy określone relacje są spełnione.

d/ Ankiety

Ankieta jest zbiorem pytań. Wszelka indagacja, śledztwo jest więc ankietą w myśl naszego rozumienia. Również egzamin jest tu ankietą. W ankietach należy rozróżnić dwa rodzaje pytań :

I/ pytania zamknięte

II/ pytania otwarte

Pytania zamknięte mają taką postać, że odpowiedzią na nie jest wybranie alternatyw z góry określonego zbioru odpowiedzi /np. tak, nie/. Pytania otwarte nie zakładają z góry ściśle ograniczonego zbioru odpowiedzi jak np. pytanie : Co myślisz o swoim szefie ?

e/ Odpowiedzi

Mam tu na myśli teksty, które otrzymuje np. jako odpowiedź na ankietę, czy też jakiegokolwiek inny zbiór pytań. Wiele tekstów można napewno traktować jako odpowiedzi /np. wiele teorii jest odpowiedziami na określone pytania/. Tekst otrzymany z przesłuchania świadka jest np. odpowiedzią w naszym sensie.

f/ Programy /plany/

Są to teksty opisujące przebieg czynności. Np. opis wykonania jakiegoś przedmiotu będzie programem, opis rozwoju jakiejś dziedziny przemysłu, kraju itp będą również planami.



g/ Sprawozdania /relacje/

Są to opisy stosunków zachodzących między pewnymi przedmiotami, też opisy pewnych zdarzeń / procesów, realizacji programów /.

h/ Wykazy

Mamy tu na myśli tego rodzaju teksty jak np. informacje bibliograficzne, wykazy etc.

Proponowany podział tekstów jest być może naiwny a napewno niekompletny. Nie jest on również rozłączony gdyż punkty b/ Teorie oraz c/ Odpowiedzi napewno w znacznym stopniu się pokrywają /może nawet całkowicie ?/. Również częściowo pokrywają się punkty f/ Sprawozdania oraz g/ Informacje.

Napewno nie należy sądziwać się wszelkimi możliwymi tekstami a z podanych grup wybrać te, które mają większe znaczenie praktyczne i w miarę możliwości są ze sobą jakos związane. W tak wybranej podklasie tekstów, mając na uwadze konkretne zastosowania można kusić się już o jakies jednolite ujęcie materiału.

2. Własności tekstów

Mając zadany tekst możemy pytać czy ma on jakas interesujaca nas własność czy też nie. Własności o które możemy pytać zależą napewno od rodzaju tekstu i oczywiście od interesujacych nas zagadnień. Tym nie mniej pewne typy pytań zależą wyłącznie od rodzaju tekstu a nie od jego treści.

a/ Teksty niesidentyfikowane

Napewno można tu pytać co najmniej o dwie rzeczy :

I/ Jakie prawidłowości formalne zachodzą w zadanym tekście /jaką ma on gramatykę/

II/ Czy dany tekst spełnia określone reguły /zadaną gramatykę/

Celem odpowiedzi na te pytania mogą być konieczne badania np. statystyczne tekstu, bądź też badania natury kombinatorycznej.

b/ Teorie

Odnośnie tekstów uzasadniających może <sup>na</sup>pytać jak się wydaje o następujące rzeczy :

- I/ Niesprzeczność tekstu - sens tych pytań jest oczywisty
- II/ Jednorodność tekstu. Mam tu na myśli pytanie czy tekst dotyczy jednej sprawy czy wielu i ewentualnie ilu ? Jakich ?
- III/ Kompletność /zupełność/. Chodzi tu o to czy zadany tekst w jakimś sensie jest wyczerpujący.
- IV/ Poprawność - tj zgodność z ustalonymi zasadami, np. w odniesieniu do tekstów dedukcyjnych - z prawami logiki.
- V/ Zwartość - czy w tekście nie ma fragmentów powtarzających się.

c/ Kodeksy

W tej klasie tekstów podstawowymi chyba są :

- I/ Niesprzeczność
- II/ Zwartość
- III/ Kompletność
- IV/ Poprawność definicji /brak błędnego koła/
- V/ Analiza norm - np. za jakie przestępstwa grozi kara od 2 do 5 lat więzienia.

d/ Ankiety

W ankietach napewno ważne są następujące zagadnienia :

- I/ Kiedy pytania otwarte można zastąpić zamkniętymi i odwrotnie ?
- II/ Czy wszystkie pytania można zamknąć ?
- III/ Czy lepiej jest zadawać pytania zamknięte czy otwarte? /zależy oczywiście do jakiego celu/
- IV/ Kompletność
- V/ Skuteczność - chodzi tu o stwierdzenie czy zadany ciąg pytań odpowiedzi dostarcza żądanej informacji
- VI/ Układanie pytań /ankiety/ celem uzyskania od pytanego określonych informacji

e/ Odpowiedzi

W tekstach tego rodzaju można zadawać pytania następujące:

- I/ Niesprzeczność



II/ Kompletność

III/ Zwartość

IV/ Jednorodność

V/ Poprawność

VI/ Mając zadaną odpowiedź możemy pytać jakie z nią jest związane pytanie ?

f/ Programy

Tu chyba trudno o pytania natury ogólnej. Myślę, że pytania odnośnie własności takich tekstów można dopiero formułować o konkretne dziedziny których te programy dotyczą.

g/ Sprawozdania

Tu można chyba zadawać pytania zbliżone jak w tekstach dotyczących teorii oraz odpowiedzi.

h/ Wykazy

Tutaj również częściowo można zadawać pytania takie jak odnośnie teorii oraz odpowiedzi.

3. Zbiory tekstów

Do tej pory podane uwagi dotyczyły pojedynczych tekstów. Wydaje się, że oddzielnej analizy wymaga rozpatrywanie zbiorów tekstów. Przez zbiór tekstów rozumiem np. zeznania świadków na temat tego samego zdarzenia, zbiór odpowiedzi na tą samą ankietę etc. Odnośnie takich zbiorów tekstów możemy pytać o czy dotyczą one tego samego tematu /są spójne/, niesprzeczne, etc. Bliższe sprecyzowanie pytań zależy oczywiście od rodzaju tekstu oraz tematyki.