



# Semantyczny model aktów prawnych - czy można okiełznać legislacyjny chaos

Michał Gałusza WAT

- Opis sytuacji

W administracji rządowej i samorządowej codziennie powstają akty prawne, które regulują procesy administracyjne, do których zarówno obywatele jak i przedsiębiorcy muszą się przystosować. Nieprzystosowanie naraża na kary a przynajmniej na stratę czasu przy wypełnianiu formalnych wymogów regulacji.

- Problem

Powstające regulacje nie mają jasnego powiązania między sobą. W jednej regulacji są zmiany w ramach różnych obszarów. Regulacje powstają na różnych szczeblach administracji w różnych momentach. Informacji o zmianach nie można obserwować na zasadzie automatycznych powiadomień. Nie można subskrybować obszarów legislacyjnych, które są szczególnie istotne np.: kancelaria podatkowa nie dostaje automatycznych powiadomień n/t interpretacji US lub zmian w przepisach wykonawczych w obszarze prawa podatkowego.

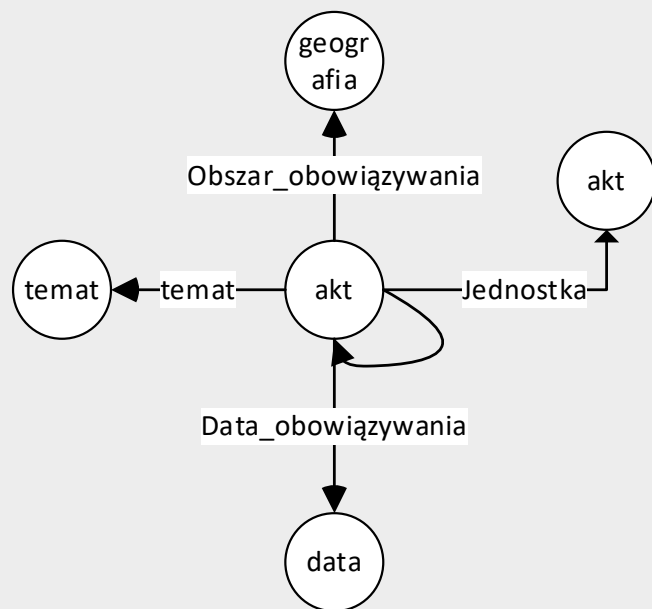
- Konsekwencje

Przystosowanie się wymaga zarówno doświadczenia (wiedzy o już istniejących regulacjach) jak również czasu aby z różnych źródeł (dokumentów) informacje połączyć oraz wymusza aktywną obserwację informacji o zmianach w regulacjach.

Semantyczny Model Aktów Prawnych (przykład) :

Dodatkowe informacje o regulacjach takie jak data obowiązywania, obszar tematyczny, wskaźnik do wersji poprzedniej, jednostka publikująca, obszar terytorialny zastosowania

Model pozwala na selekcję aktów prawnych pod różnym kątem metadanych, których szukamy oraz pozwala na wyznaczenie zmian przepisów w stosunku do poprzednio obowiązujących.

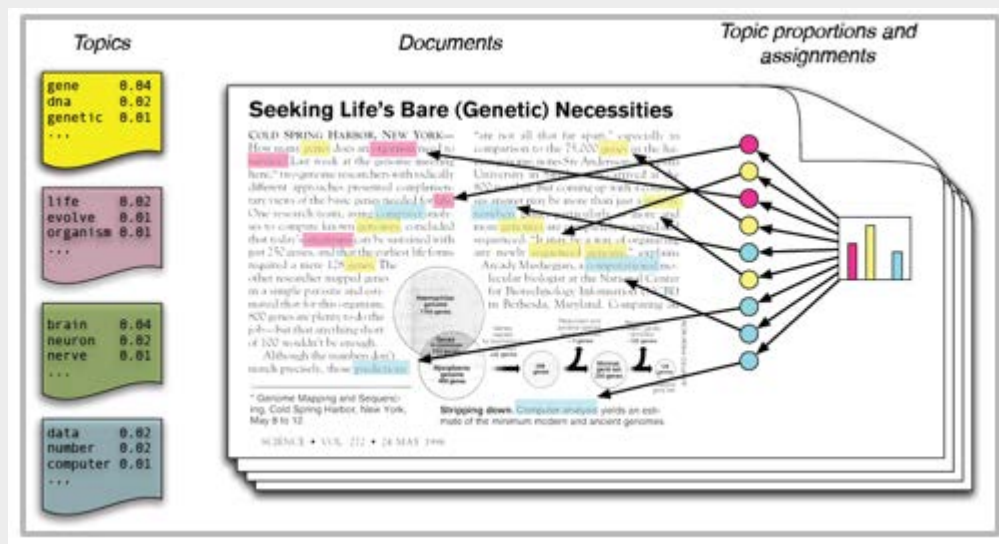


Model grafowy aktu prawnego

Metadane aktu możemy podzielić na kategorie:

- Szttywne i ustalane w momencie tworzenia (jednostka tworząca, okres obowiązywania, poprzednia wersja)
- Charakterystyczne dla treści aktu

Modelowanie metadanych charakterystycznych dla treści jest zadaniem z kategorii „topic modeling” tzn przypisaniem klasy lub klas semantycznych (tematu) dla dokumentu lub jego części w oparciu charakterystyki treści.



## Hashtagi.

### Ręczne nadanie hashtagu aktowi lub jego części.

Wady:

1. Centralnie zarządzany zbiór hashtagów.
2. Ręczna i nieidealna klasyfikacja aktów – pomijanie hashtagów lub błędne przypisania.
3. Brak lub sztywna hierarchia hashtagów.
4. Przy zmianie korpusu dokumentów lub struktury hashtagów praca ręczna musi być potwierzona.

## Detekcja tematów dokumentów

Istnieją już algorytmy detekcji tematów dokumentów lub ich sekcji w oparciu o rozkład statystyczny słów:

Probabilistic Latent Semantic Analysis - PLSA  
Latent Dirichlet Allocation - LDA

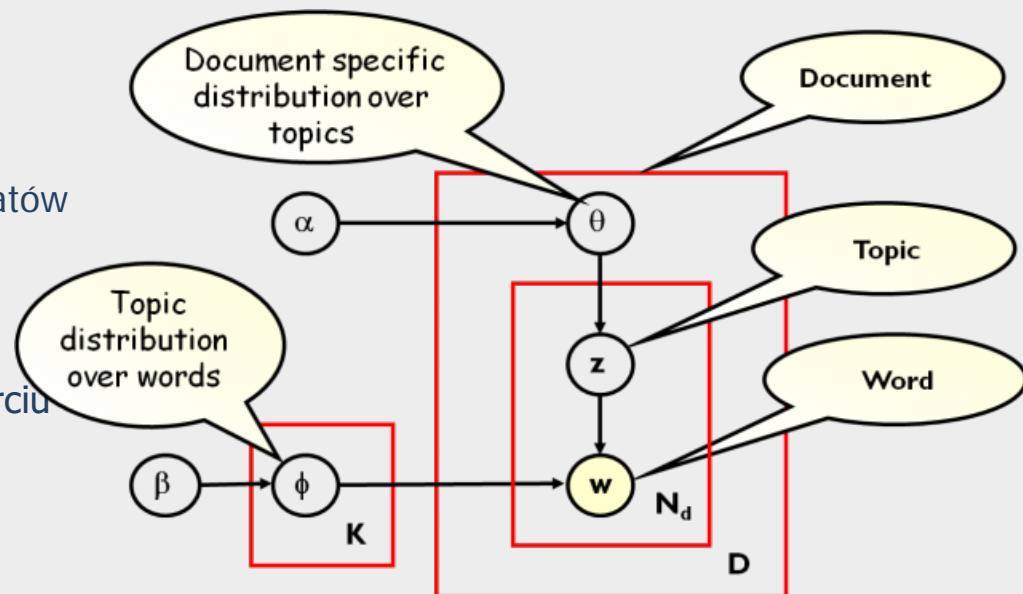
Są to algorytmy klasy uczenia bez nadzoru.

Zalety:

1. Adekwatne wybieranie tematów w zależności od treści
2. Automatyczna detekcja tematów.
3. Automatyczne przypisanie nowego dokumentu do tematów

Wady:

1. Brak możliwości zdefiniowania hierarchii tematów.
2. Podanie ilości tematów i ich późniejsze opisanie w oparciu o słowa, które zostały do nich zaklasyfikowane.



## Detekcja hierarchii tematycznych

Rozszerzenie modeli detekcji tematów – wyznaczanie drzewa tematów adekwatnych dla dokumentu. Struktura drzewa jest wyznaczana na podstawie danych. Obecnie najnowszy obszar badań zakresie modelowanie tematów.

Algorytm: hLDA – Hierarchical Latent Dirchlett Allocation

Zalety:

1. Adekwatne wybieranie tematów i ich struktury zależności od treści
2. Automatyczna detekcja tematów.
3. Automatyczne przypisanie nowego dokumentu do struktury

Wady:

1. Na obecnym etapie nie można ustalić liczby tematów ani struktury drzewa

