

Sztuczna Inteligencja

IBM WATSON

Włodzisław Duch

Katedra Informatyki Stosowanej UMK

Google: Włodzisław Duch

Watson Q/A

- System QA firmy IBM. Język to kolejne wielkie wyzwanie po szachach. Projekt rozpoczęty w 2007 roku. Watson tutorials.
- “Massively parallel probabilistic evidence-based architecture”.
- Początkowo odpowiedź wymagała 3 godzin, skrócono ją do 3 sek.
- Hardware: stacje robocze Power7 (10 Power7 server blades)+90 serwerów, 4 procesory/serwer, 8 rdzeni/procesor (w sumie 2880 rdzeni), 15 TB RAM, 20 TB dysk, 80 TFlop – duże wymagania.
- Wejście czysto tekstowe, brak analizy mowy czy obrazu.
- W 2011 IBM Watson wygrywa z dwoma mistrzami teleturnieju Jeopardy (Va Banque). W konkursie ramię robota naciskało przycisk.
- Zaczęła się era “kognitywnych obliczeń”, od zabawy do zastosowań medycznych i AI dla biznesu.

PIENSE

THINK

ΣΚΕΨΟΥ

DENKE

ΣΙΝΑΤΙΝΙΣ

PENSER

\$200

\$4,000

\$600

Ken

WATSON

BRAD

Maxwell's silver hammer

FRANK SINATRA

Brown



96%



11%



7%



Watson

- Podstawą jest otwarty framework IBM UIMA: unstructured information management application; nadaj strukturę informacjom.
- Wszystkie zasoby tekstowe zostały zindexowane wcześniej, nie używano szukania w czasie gry.
- Użyto wielu różnych źródeł informacji.
 - Podstawą była Wikipedia, Project Gutenberg, tezaury, doniesienia gazetowe itp.
 - Dodatkowe informacje aktywnie szukane w sieci i dodane do korpusu treningowego.
 - Tekst bez anotacji i bez struktury.
 - Wykorzystano analizę struktury Wikipedii z projektu Yago, oraz analizę pojęć z Wordnet.
- Jedynie 2% odpowiedzi w Jeopardy! było bezpośrednio dostępne w utworzonym korpusie, pozostałe wymagały wnioskowania.
- Klucz: integracja różnych źródeł informacji i technik wnioskowania.

Składowe DeepQA

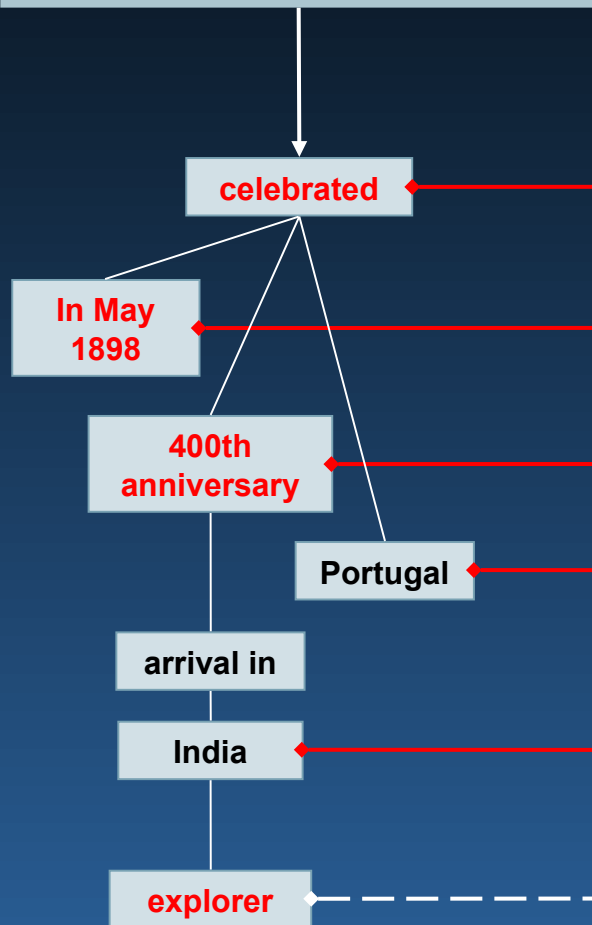
Okolo 100 różnych technik.

- Akwizycja: utworzenie korpusu (offline), zabawa w gry słowne.
- Preprocessing
- Liczne narzędzia NLP, parsery
- Wyszukiwanie odpowiedzi, mechanizmy wnioskujące
- Oceny jakości odpowiedzi.
- Strategie gry
- Strategia wciskania przycisku po osiągnięciu pewności odpowiedzi.

Złożone pytania

Question:

In May 1898 Portugal celebrated the 400th anniversary of this explorer's arrival in India

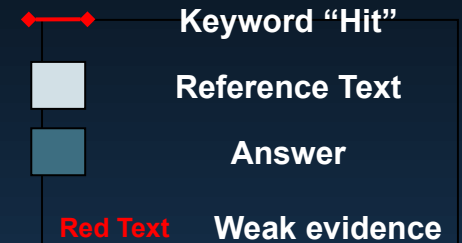


Supporting Evidence:

In May, Gary arrived in India after he celebrated his anniversary in Portugal



Legend



Hipoteza:
odpowiedzią jest "Gary" ale system nie może się opierać tylko na dopasowaniu słów kluczowych bo to słabe potwierdzenie.



Watson używa głębszej analizy

Question:

In May 1898 Portugal celebrated the 400th anniversary of this explorer's arrival in India.

celebrated

Portugal

May 1898

400th anniversary

arrival
in

India

explorer

Supporting Evidence:

On the 27th of May 1498, Vasco da Gama landed in Kappad Beach

landed in

27th May 1498

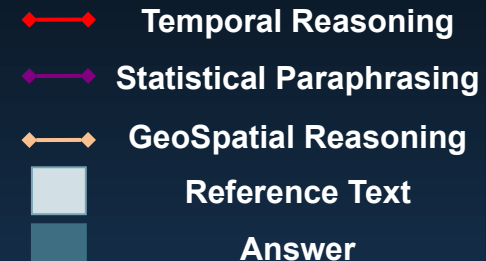
Date
Match

Para-
phrases

Geo-
KB

Kappad Beach

Vasco da
Gama



Trudniej to znaleźć

- Szersze szukanie
- Eksploracja wielu hipotez
- Ocena potwierdzenia
- Wiele algorytmów wnioskujących

Preprocessing

- Określenie kategorii pytań
 - faktoidy
 - rozkładalne na dwa lub więcej członów
 - puzzle
- Określenie kategorii odpowiedzi (lexical answer type, LAT)
 - film? osoba? miejsce? piosenka? nowela?

Około 2500 z 20,000 pytań należało do tych kategorii.

Analiza NLP Watsona

- Parsery gramatyczne
- Dekompozycja na prostsze fragmenty
- Tagi semantyczne składowych pytań
- Rozstrzygnięcie ko-referencji
- Odkrywanie relacji między słowami
- Rozpoznanie nazw własnych (named entity)

Szukanie przydatnych tekstów

- Wyspecjalizowane szukarki: Indri (projekt Lemur), Apache Lucene, SPARQL RDF Query
- Szukanie składowych pytań podobnych do użytych w wyszukiwaniu informacji z sieci
- Skupienie na odzyskaniu (recall) wszystkiego, niska czułość.
- Generowanie setek hipotez
- Dla dobrze określonych kategorii LATs (np. nazwy krajów, stanów USA, prezydentów) lista możliwych odpowiedzi generowana jest bezpośrednio.
- Dla pozostałych konieczne jest wnioskowanie

Ocena hipotez

- Ocena wygenerowanych hipotez przez
 - miękkie filtrowanie (soft filtering), szybkie algorytmy zostawiają około 100 hipotez
 - szukanie wsparcia dla każdej z hipotez, dodatkowe zapytania skierowane do korpusu
- Oceny odpowiedzi
 - ponad 50 różnych algorytmów
 - od prostego zliczania słów do złożonych form rozumowania uwzględniającego relacje czasowe i przestrzenne
 - tworzony jest profil potwierdzeń (evidence profile): zaufanie do źródeł, taksonomia, geoprzestrzenne i czasowe relacje ...
- Łączenie odpowiedzi
- Końcowy ranking i estymacja wiarygodności.

Strategie końcowe gry

- Kiedy się zgłaszać do odpowiedzi (naciśnięcie przycisku)?
 - Ocena pewności $>50\%$
 - Ocena ryzyka uwzględniająca uzyskaną przewagę przy niższych wartościach.
 - Nie zgłasza się jeśli błąd może doprowadzić do przegranej.
- Ile i jak warto deklarować?

IBM
Research
Project
(2006 –)

Jeopardy!
Grand
Challenge
(Feb 2011)

Watson
for
Healthcare
(Aug 2011)

Watson
for Financial
Services
(Mar 2012)

Watson
Ecosystem
(2014–)



R&
D

Demo

Komercja

Ekspansja

Liczne
zastosowania





23 mln

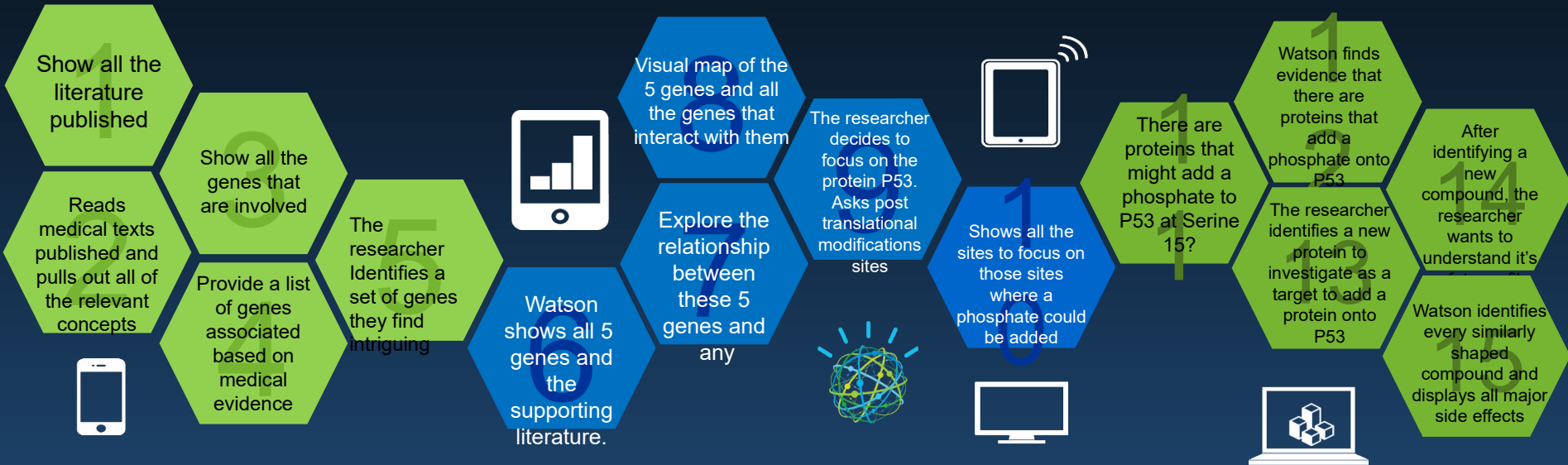
Liczba artykułów,
codziennie
dodawanych, w oparciu
o które podejmowane
są decyzje medyczne.

Watson: Transforming the Discovery of Breakthrough Treatments

Gain Awareness

Increase Understanding

Clarify and Validate



Watson Enhancements



Literatura

- Najwięcej informacji: <https://www.ibm.com/watson>
- Prezentacja [algorytmów IBM Watson](#) (ITT Bombay)
- Szczegółowy opis: [AIMagazine-Building Watson-Overview of DeepQA Project](#)
- Końcówka gry Jeopardy! https://www.youtube.com/watch?v=WFR3lOm_xhE
- [IBM Debater i Watson Discovery](#)

Project Debater is the first AI system that can debate humans on complex topics, help people to build persuasive arguments and make well-informed decisions.

- Obecnie IBM połączył symboliczne możliwości Watsona z dużymi modelami językowymi, oferując system WatsonX dla dużych firm.