

IBM Research - Tokyo

IBM 👸

2011年2月14-16日、IBMリサーチが開発した質問応答システムWatsonが米国の人気クイズ番組「Jeopardy!」で勝利しました

"Jeopardy!" とは

- **1964**年から続くアメリカのクイ ズ番組
- ■月曜〜金曜の夕方、既に**9000**回 以上放映
- 歴史・科学・スポーツ等幅広い 分野から出題
 - 知識が問われる正統派のクイズ
 - 言葉遊び的なパズルも

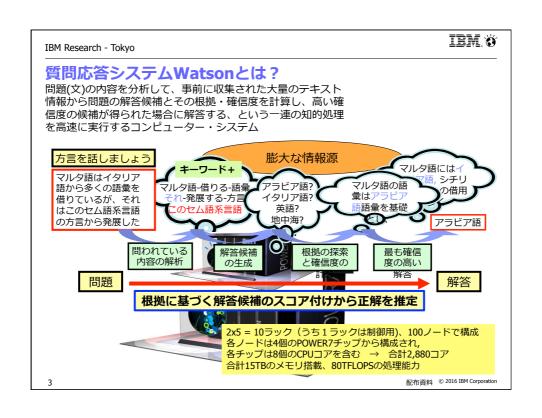


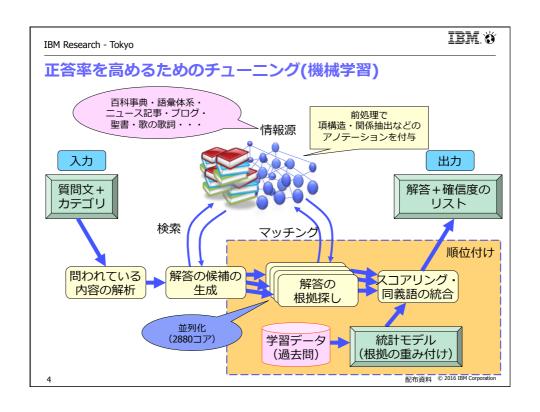
- ■解答者3人が獲得金額を競う形式(Watson+2名の対戦相手)
 - Ken Jennings氏(最多連勝記録保持者)
 - **Brad Rutter**氏(累積最高賞金記録保持者) 正答率:

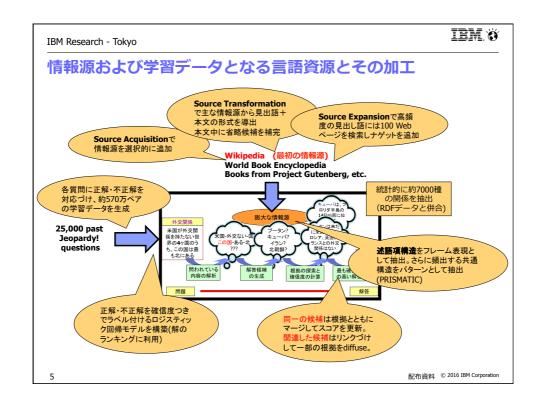
Jennings氏: 88%, Rutter氏: 89%, Watson: 88%

- ・ 「質問応答」というテキスト情報に基づく解答の計算手法が広く認知された
- ITでは解決できないと思われていた問題をIBMにご相談いただくようになった
- · Watsonで解決可能な問題の拡大(Cognitive Computingとして一般化)

配布資料 © 2016 IBM Corporation







IBM Research - Tokyo

IRM Ö

言語資源の存在を前提とした知的システムの設計

▶ タスク設計

利用可能な言語資源の質・量と現状の自然言語処理や機械学習などの精度でも実用性のあるタスク

> 既存言語資源の有効な利用

- 意図したタスク専用に構築された言語資源は見つからない
- 新たなタスク専用の言語資源を独自開発するコスト・期間は大きな負担
- 言語資源を加工することで必要とされる情報を抽出・導出
- さらに抽出された情報を統計的に集約することで汎用パターンやルール を獲得
- Word embeddingなどの手法で一部同義語辞書などを代替可能に

> ルールベース手法と機械学習・統計的手法の併用

高precision/低recall + 低precision/高recall をバランス 最終的にはルールベース手法を逐次改良して高precision/高recallを達成

▶ 当初の精度を維持するために保守作業が不可避・不可欠

情報源や学習データの更新、タスクの入力となるデータの変化に対応

商業的には保守作業のルーチン化・簡素化が鍵となる

6 配布資料 © 2016 IBM Corporation

IBM Research - Tokyo

IRM S

Watsonの質問応答技術とその事業化

質問応答技術を起点に、ビッグデータ基盤構築からバートナーソリューションまで幅広く支援

医療保険承認業務

創薬支援

コンタクトセンター支援

資産運用

医療機関からの治療事前承認 大量の文献、特許、化合 業務を省力化、かつエビデン 物などの情報から創薬に スに基づく医療にドライブ

物などの情報から創薬に コンタクトセンターの回答支援 ファイナンシャルアド つながるヒントを提示 スマートフォン経由の自動応答 バイザー業務を支援

医療教育分野

質問応答+既存技術を統合しブランド化

医師国家試験を題材に、問題 発見解決型教育ツールを開発

パートナーソリューション

IBM Watsonプラットフォームを提供し、ビジネスパートナーによる ソリューション構築・提供を支援

医療分野

患者の症状に対し、診断の参 考になる情報提供、必要な検 査、良いアウトカムにつなが る治療方法の研究を支援



クラウド + SoftLayer/IBM Bluemixを利用したIBM Watson プラットフォーム構築



配布資料 © 2016 IBM Corporation



IRM.

スケア/ライフサイエンス分野の言語資源とその役割

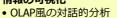


データベース中の要素や 実験対象についての

知見・コンテキストを獲得

Bioinformatics ポータルサイト テキストマイニング







傾向/パターン/相関の分析

情報抽出 • 用語/実体/関係の抽出



•1000ドル・ゲノム分析 •大規模データベースとの統合

電子カルテ 臨床試験・副作用 診療ガイドライン等



MEDLINE (約2500万文書)やライフサイエンス特許文書など

• テキスト情報とカタログ属性による情報の記述

MeSH シソーラス (約2.7万語のdescriptor)やIPC(International Patent Code) • 生医学分野の主題の階層的な体系

PubMed等の 検索サービス

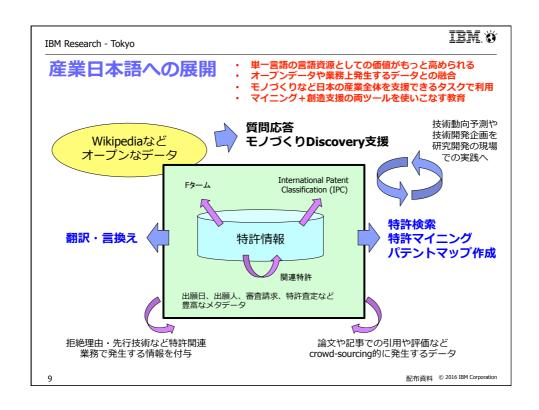
UMLS メタシソーラス (280万概念 + 1120万語の概念名称)やGOなど • 生医学分野の多様な情報源に現れる専門用語の体系化

米国医学図書館などにより公開されたデータ

日本語の言語リソースは貴重:臨床医学オントロジーなど

8

配布資料 © 2016 IBM Corporation



IBM Research - Tokyo

ありがとうございました

IBM、IBMロゴ、Smarter Planetアイコンおよび POWER7は、世界の多くの国で登録されたInternational Business Machines Corp.の障標です。他の製品名およびサービス名等は、それぞれIBMまたは各社の商標である場合があります。
現時点でのIBMの商標リストについては、http://www.ibm.com/legal/copytrade.shtmlをご覧下さい。

配布資料 © 2016 IBM Corporation