Watson Explorerを使ったテキストマイニング

株式会社エクサ
テクノロジーイノベーション部
堀　扶
tasuku-hori@exa-corp.co.jp

目次

- はじめに
- 概要
- 試行
- まとめ

©2016 EXA CORPORATION

コンピュータが仕事を奪う日

2016年6月某日
1F受付が机ごと消える

代わりに設置されたのがコレ

©2016 EXA CORPORATION

きっかけ

- 『人工知能』『コグニティブ』
  という世間の声

- 『Watson をやりたい』
  という社長の声

- 『これからはWatson Explorerが中心になりそう』
  という案件経験者の声

©2016 EXA CORPORATION

Watson Explorer

- テキスト検索
  - 様々な検索元に対応
  - Twitterをはじめとする様々なWebコンテンツから社内の非定型文書まで
- テキスト分析
  - 多くのの言語に対応した分析が可能

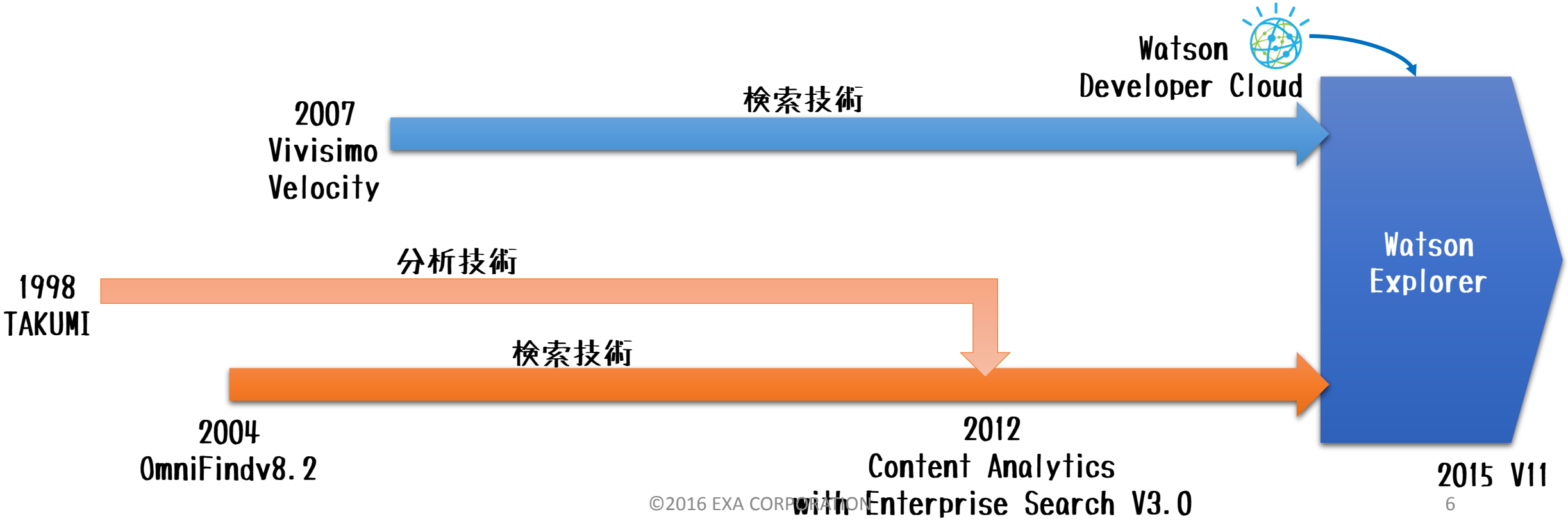
- Watson Developer Cloud

※75日間限定ライセンス

©2016 EXA CORPORATION

Watson Explorer(WEX): 2つの製品を融合

- Watson Explorer Foundational Components(Vivisimo)
- Watson Explorer Analytical Components (OmniFind/TAKUMI)



©2016 EXA CORPORATION

WEX Foundational Components(Vivisimo)

- IBMが買収したVivisimo社は「エンタープライズサーチ」の会社
  - 買収後InfoSphereというブランドで販売
- 検索機能のカスタマイズ範囲がOmniFindより多い
  - 自動でインデックスを作成してくれる
  - Fast Indexing：指定した項目で検索できる
  - オートコンプリート機能：ユーザ独自の辞書を作成できる
  - メタデータ活用：XPathで切り出した項目で検索できる
  - スポットライト機能：特定のキーワード検索された場合のみ表示可能に
- Application Builderによる360度ビューアプリの構築が可能
  - 360度ビューによる、切り口を固定した分析は可能

2012 All Edition/x86-64
2012 R2 All Edition/x86-64

RHEL Server6/x86-64
RHEL Server7/x86-64

Watson APIをコールできる

©2016 EXA CORPORATION

WEX Analytical Components (OmniFind)

- 検索機能と分析機能を併せ持つ
  - テキストマイニングは基本OmniFind側を使う
  - 分析対象の切り口はVivisimoより柔軟性がある
- 日本語文書も分析できる
  - 分析機能のベース(TAKMI)は和光で開発
  - UIも比較的日本人向けの設計
- ソリューションパッケージ追加可能
  - エンタープライズ検索コレクション
    - ソーシャル検索、ドメイン適応検索*
  - コンテンツ分析コレクション
    - AlchemyLanguageAPI*、AlchemyVisionAPI、RDFエンティティ抽出,ソーシャル分析、市場分析*、自動車の早期段階での品質障害*、小売店VoC分析*、電気通信事業者VoC分析*

*英語のみ

2012 All Edition/x86-64
2012 R2 All Edition/x86-64

RHEL Server7/x86-64/p/z
RHEL Server6.4/x86-64/z

SLES 12/x86-64/z
SLES 11/x86-64/z

6.1/p
7.1/p

©2016 EXA CORPORATION

テキストマイニング

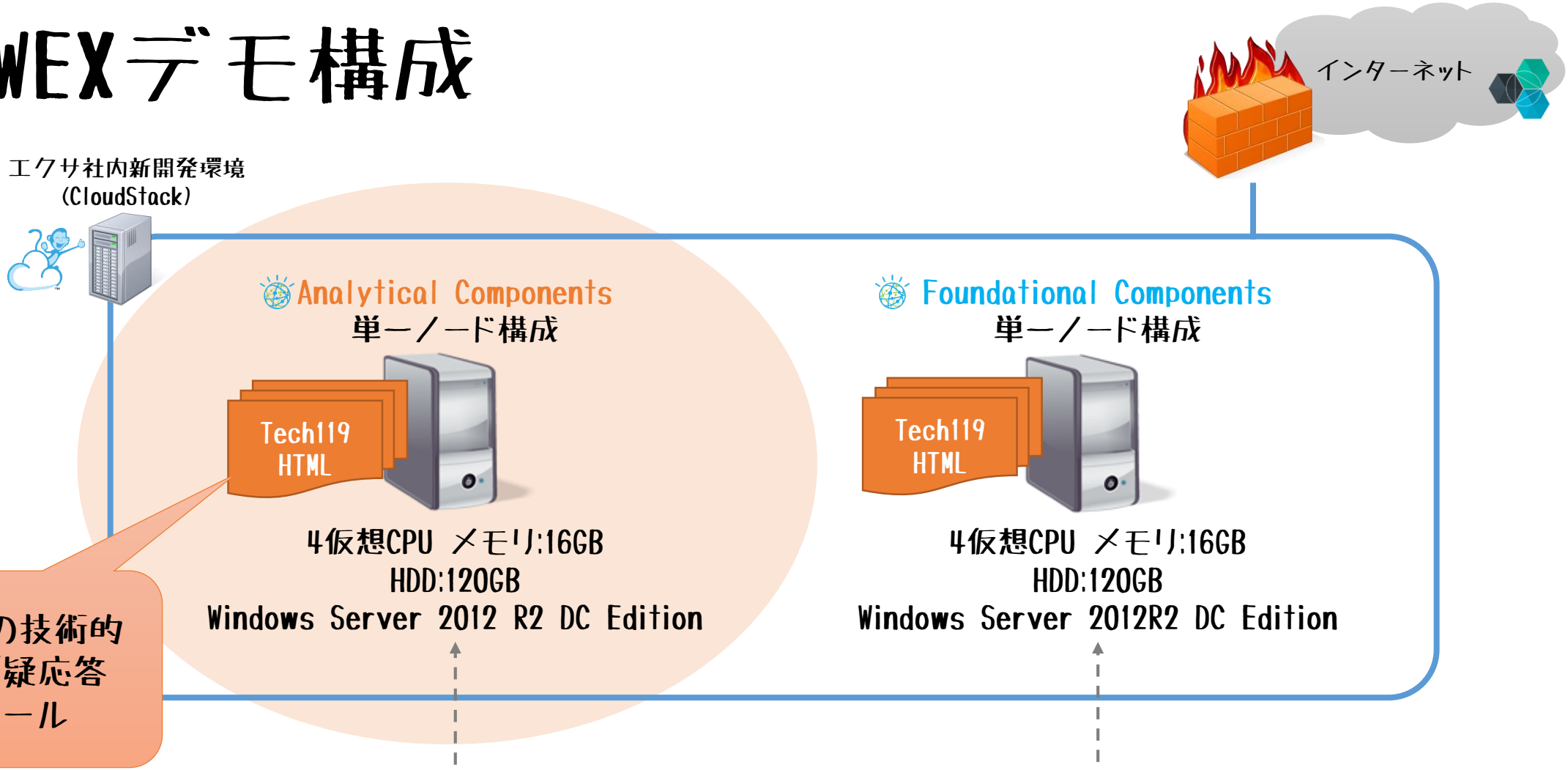
- 非構造化データから新たな法則性や傾向を導き出すこと
  - 非構造化データ：文章などの関係モデルで定義できないデータ
- 実は方法論や手法はない
  - 非構造データの形態素解析やN-gramモデルなどの高速検索技術以外明確な手法がない
  - 業種特化で分析
- 大まかな手法はある（多くが多変量解析の応用）

- 対象データの収集
- 単語の分割、同義語の統一や専門用語の辞書化
- KKD(カン・経験・度胸)による仮説を定義
- 仮説を元に複数の切り口を定義、組み合わせて統計化
- 法則性が見つかるまで適当な手順に戻って繰り返す

©2016 EXA CORPORATION

試行

WEXデモ構成



テキストマイニングであるため
Analytical Components／ード側を使用

©2016 EXA CORPORATION

試行

WEXデモ

©2016 EXA CORPORATION

IBMは何故”Watson”とつけたのか？

- Watson Explorerに搭載されている人工知能はAlchemyのみ
  - Analytical Componentの好意・非好意分析
  - Analytical Componentの英語版のみ一部ソリューションパッケージ
  - Alchemy は MITライセンスのOpen Sourceバージョンも存在する
- Watson Developer Cloudライセンス付きだから？
  - そもそも75日間のみ

©2016 EXA CORPORATION

IBMはWEXを2つのEditionで販売

- Enterprise Edition
  - Foundational Componentsのみ
  - $275K

テキストマイニングをしたいなら、
Advanced Editionを買わなくてはならない。

- Advanced Edition
  - Foundational Components + Analytical Components
  - $500K + 対象データ容量ごと料金(分析の場合は検索の約10倍)

まさか往年の
ガンプラの抱き合わせ商法？！

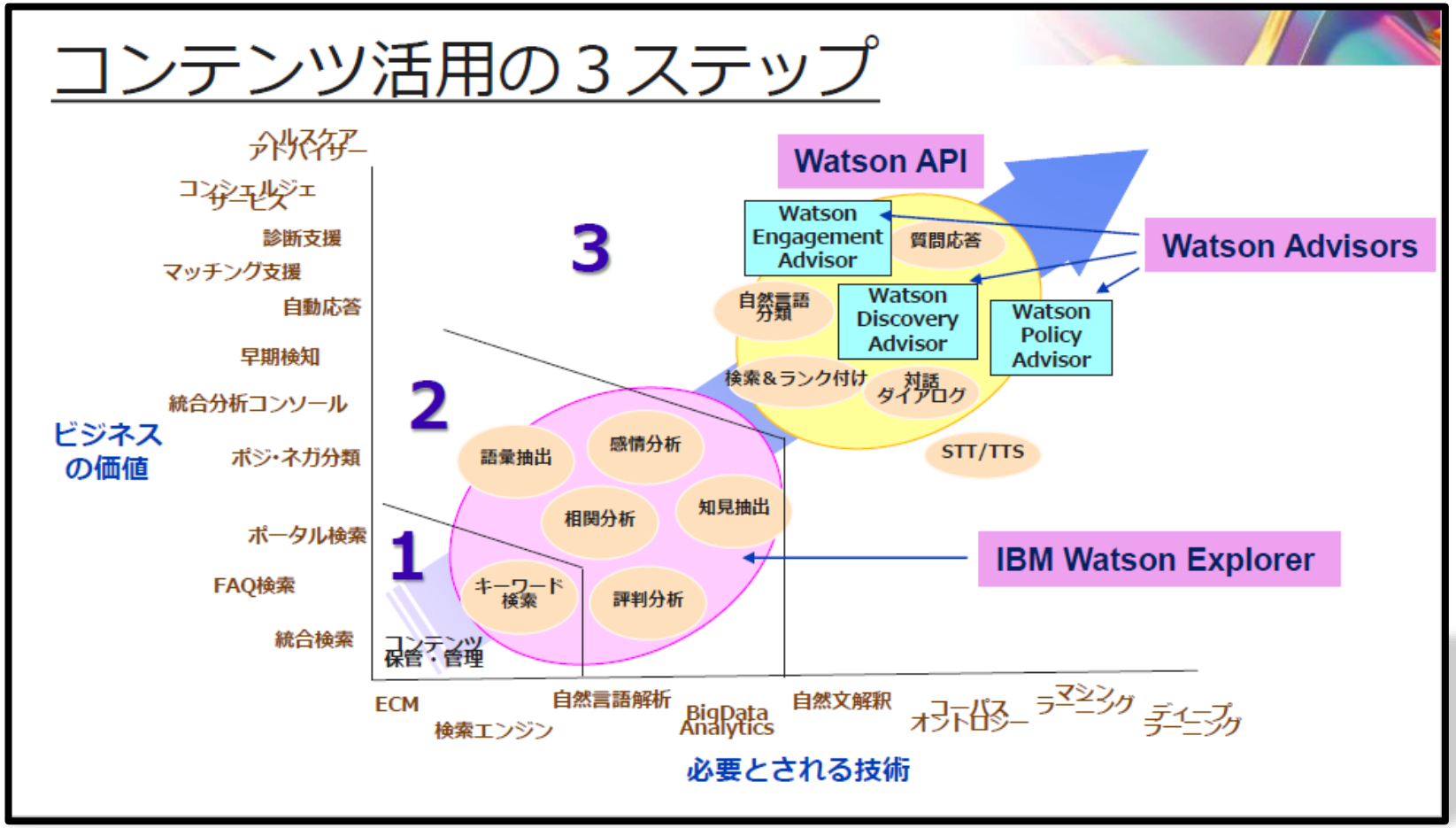
©2016 EXA CORPORATION

実は既にイベント内で語られてました..

まとめ

IBMの目論み

企業に人工知能を使ってほしい！
企業にコグニティブになってほしい！



コグニティブでないけど
SoEの活用から

1. まず、検索
2. 次に、分析
3. 当たりがついたらWatson!

よりコグニティブに！

IBM Watson Summit2016 A3-1セッション資料より抜粋

©2016 EXA CORPORATION

“Watson”を冠するIBMプロダクト

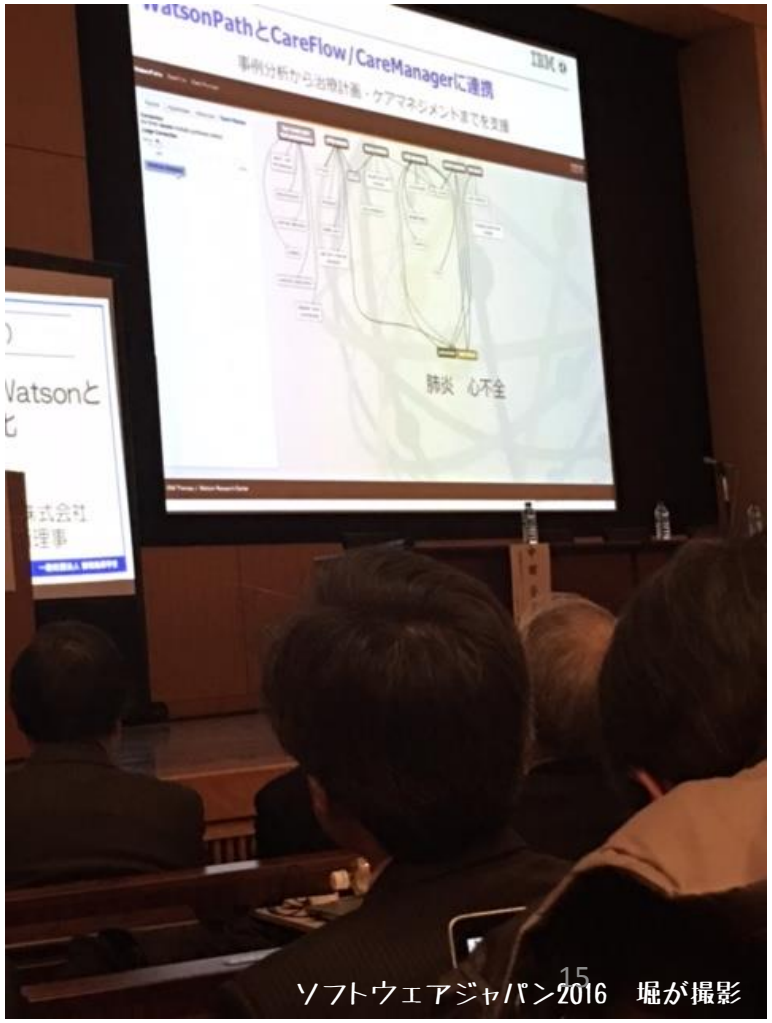
サイト「IBM Watson」に記載されているプロダクト

汎用的

用途特化

- Watson Developer Cloud
  - Watson API、Bluemix上のサービスにもなっている
  - 日本語化および販売はソフトバンク社より提供
- Watson Analytics
  - 定形データ中心の分析SaaS、元ネタはCognos
- Watson Explorer
- Watson Advisors
  - Engagement Advisor(コンシュルジュ), Discovery Advisor(仮説検証)..
- Watson Knowledge Studio
  - 推論付きの知識ベース
- Watson for Oncology
  - 癌専門医支援（症状から病名予測）
- Watson for Clinical Trial Matching
  - 癌臨床試験適合の可能性を識別

※おそらく個別要件に合わせてWatsonをカスタマイズ・学習させて提供



ソフトウェアジャパン2016　堀が撮影

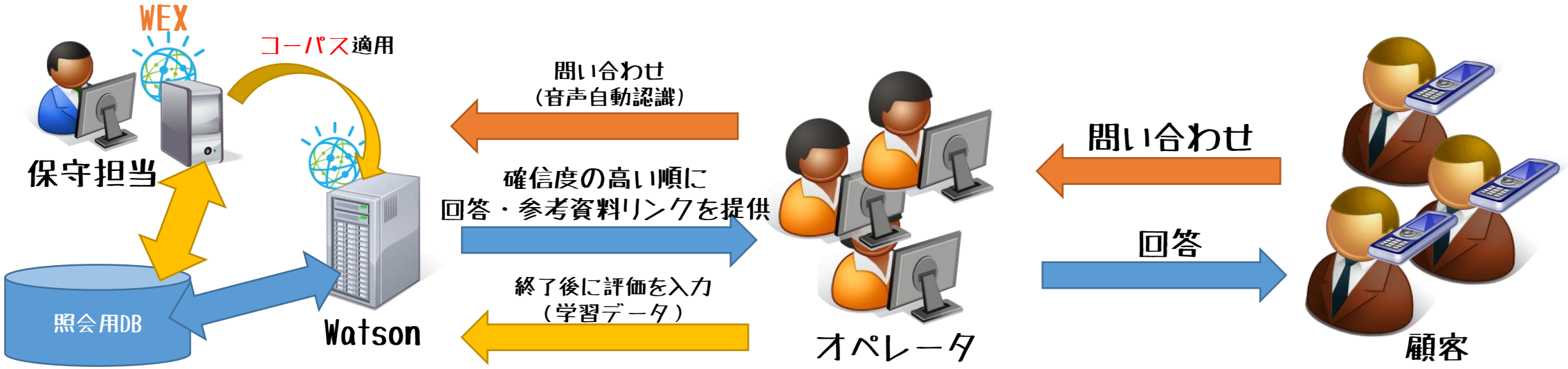
Watson IoT Platformなどサイト上に紹介のないプロダクトも..

©2016 EXA CORPORATION

では、IBMは何故WEXを押していたのか

推測①『コーパス』作成にWEXを

- 金融業コールセンタプロジェクトでは実際にWEXも使用



- 回答率を上げるためのコーパス（学習データ）作成が難しい
  - 試行を繰り返してよりよいコーパスを導き出す必要がある
  - 照会用DBや社内情報などからより良いコーパスを作成、適用する
    - WEXを使ってコーパス作成を支援

©2016 EXA CORPORATION

では、IBMは何故WEXを押していたのか

推測②回答率の高い顧客応対基盤へ
コールセンタ

- Analytical Componentsのソリューションパッケージ化
  - 業種特化したソリューションパッケージとして別売りするのでは
  - Watson APIではなく個別実装するかもしれない
    - ソフトバンク社のリソースとは別なのでIBM側のカスタマイズ自由度が高い

- Foundational Componentsの360°ビュー機能をコールセンタメンバ向け画面として適用
  - 360°ビューによるマッシュアップ開発
  - Watson Developer Cloud を継続利用するユーザになってもらい、毎月売上が立つお客様に..

©2016 EXA CORPORATION

まとめ

Mac Items - IBM Watso

watson.webpot.local:8080/AppBuilder/mac_items

開発 本番 FC FC KC FC RM FC-AB FC-AB Admin AC AC分析 AC3 AC3分析

IBM Watson Explorer

search

Home

【参考】360° ビュー （1/5）

Edit page

Edit

マクドナルド商品一覧

filter

6 of 21 columns

商品名	重量 (g)	ビタミンB1 (mg)	ビタミンB2 (mg)	ビタミンC (mg)	食物繊維 (g)
1971炙り醤油ジャパン	243	0.2	0.3	11	2.3
	246	0.2	0.23	12	2.6
	174	0.26	0.13	1	2
	207	0.19	0.13		
	187	0.16	0.3		
	164	0.26	0.13		
ダブルチーズバーガー	178	0.12	0.29		
		0.19		0.11	

1971炙り醤油ジャパン

製品重量 g/個 243

エネルギー Kcal 590

炭水化物 g 35.6

Follow

※日本マクドナルド 公式サイトより
データおよび一部画像を引用

watson.webpot.local:8080/AppBuilder/mac_items/536bca5ca6336fdfb97a478a878b1d81

©2016 EXA CORPORATION

まとめ

【参考】360° ビュー (2/5)

items?key-val=1 - IBM

watson.webpot.local:8080/AppBuilder/mac_items/536bca5ca6336fdfb97a478a878b1d81

開発 本番 FC FC KC FC RM FC-AB FC-AB Admin AC AC分析 AC3 AC3分析

IBM Watson Explorer

All search

Home

Edit page

Edit

商品イメージ

デフォルト提供されるWidget

1971炭り醤油ジャパン

製品重量 g/個 243

エネルギー Kcal 590

炭水化物 g 35.6

脂質 g 35.6

食塩相当量 g 35.6

デフォルト提供されるWidget

Latest Activity

Enter your comment

Post

All Comments Updates

filter feed

data-explorer-admin
ジャパーーーン！
10 MINUTES AGO in items?key-val=1

カロリーが先か?腹持ち

世界の腹時計

July 6th 2016
7:18:41 am
GMT

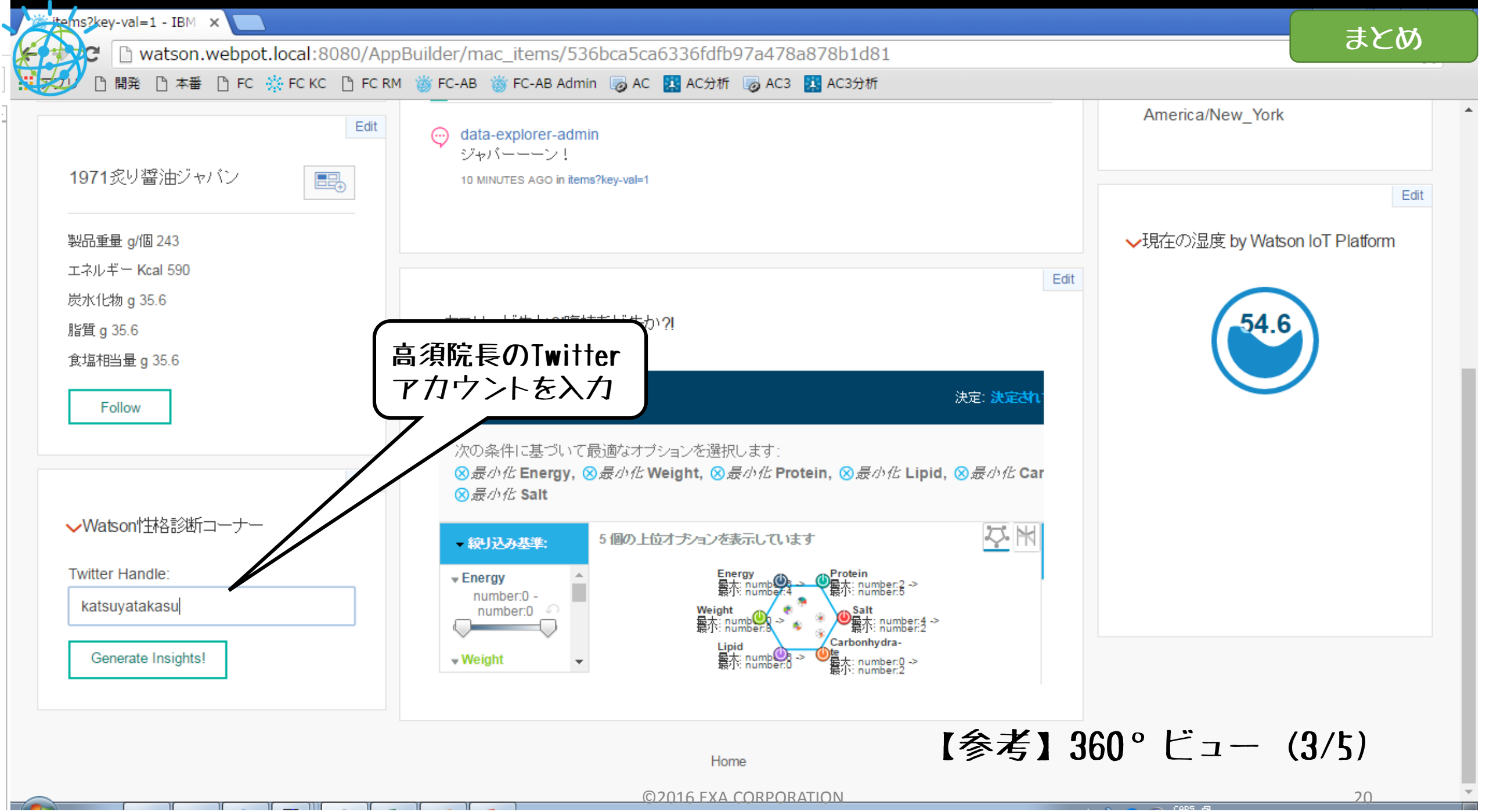
July 6th 2016
4:18:41 pm
Asia/Tokyo

July 6th 2016
3:18:41 am
America/New_York

現在の湿度 by Watson IoT Platform

54.6

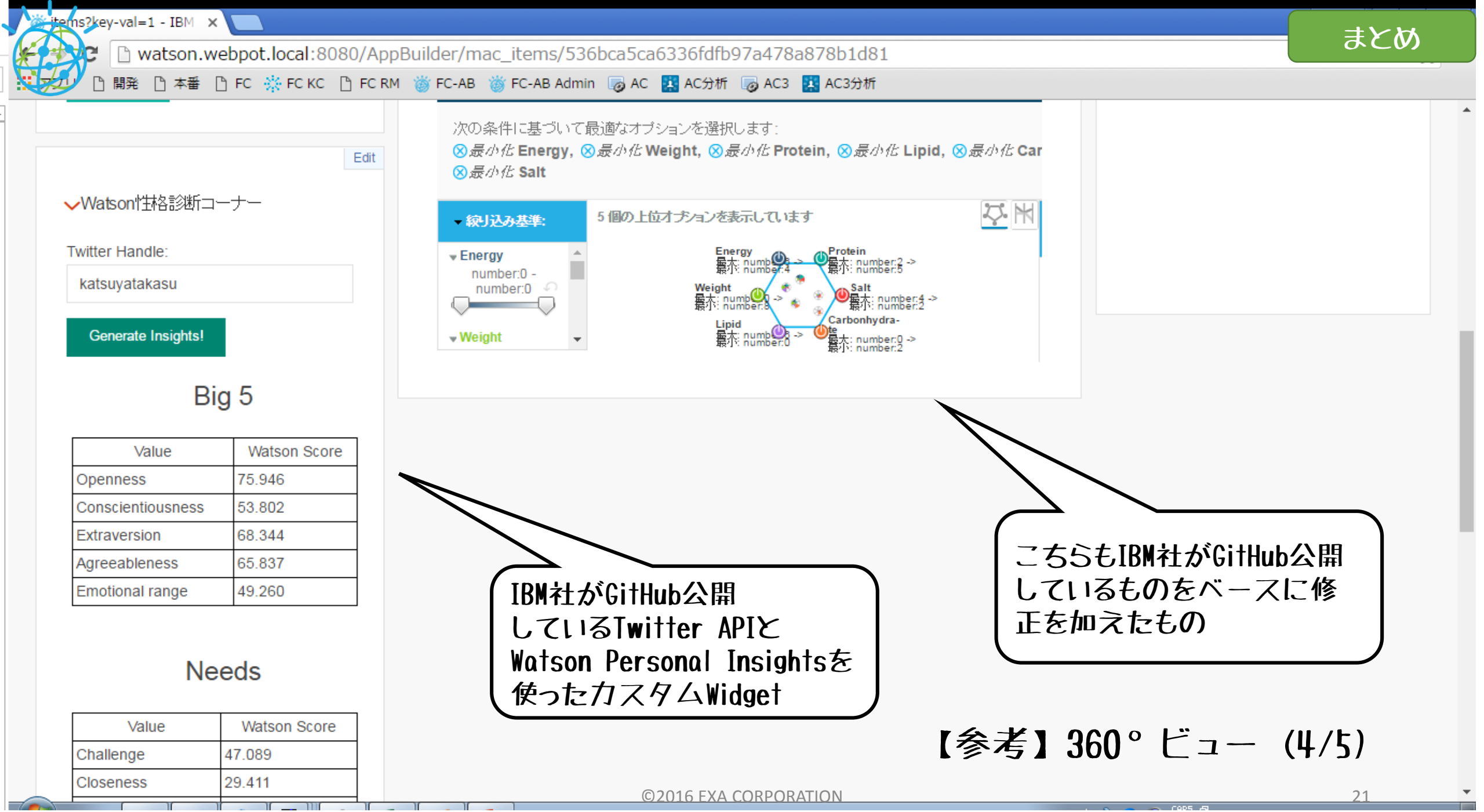
去年のIoTデモの流用
ArduinoYUNのセンサデータを表示



高須院長のTwitter
アカウントを入力

【参考】360°ビュー (3/5)

©2016 EXA CORPORATION



IBM社がGitHub公開しているTwitter APIとWatson Personal Insightsを使ったカスタムWidget

こちらもIBM社がGitHub公開しているものをベースに修正を加えたもの

【参考】360°ビュー (4/5)

©2016 EXA CORPORATION

【参考】 360° ビュー (5/5)

TradeoffAnalytics.js V1を使用。V2は画面遷移があり2画面目でXSS問題が出るため現在動作せず。

©2016 EXA CORPORATION

Salesforce.comもAI企業を買収

https://www.salesforce.com/blog/2016/06/ai-salesforce-customer-success-platform.html

Blog

The AI-First Salesforce Customer Success Platform

Jun 07, 2016 By Sarah Franklin in Apps, Cloud, Innovation, IoT Cloud, IT, Developer

Tweet Share 923 Like 324

For many visionaries and technology firms, artificial intelligence (AI) is the future. But, today at the TrailheaDX developer conference in San Francisco, Salesforce brought it into the "now," showing off its AI-first Salesforce Customer Success Platform.

With advances in machine learning, deep learning and other AI technologies, the technology industry is moving from systems-of-record and systems-of-engagement, to systems-of-intelligence to help our customers better connect with

SalesforceもCRM基盤にAIを搭載しようとしている？！

http://japan.zdnet.com/article/35080670/

Japan Edition ▼ | Blog | ホワイトペーパー | 企業情報センター ▼ | 話題の1本 | builder by ZDNet Japan

ZDNet Japan

編集部からのお知らせ
- 7/13 情報漏えい対策セミナー 年金機構から1年
- 海外ニュース記事ライター（フリーランス）募集

Google Cloud Platform　Googleが提供する開発者のためのクラウドプラットフォーム

CIO | クラウド | モバイル PC | セキュリティ | ソフトウェア | サーバ | ストレージ | ネットワーク | ミドルウェア | 運用 | マーケティング | ビッグデータ | キャリア

ZDNet Japan > 経営

セールスフォース、人工知能関連の新興企業MetaMindを買収

Natalie Gagliordi （ZDNet.com） 翻訳校正： 編集部 2016年04月05日 12時23分

いいね！ 149　ツイート　G+1 3　B! 5　Pocket 27

Salesforce.comが人工知能（AI）関連の新興企業MetaMindを買収したことが米国時間4月4日に発表された。買収の条件は明らかにされていない。

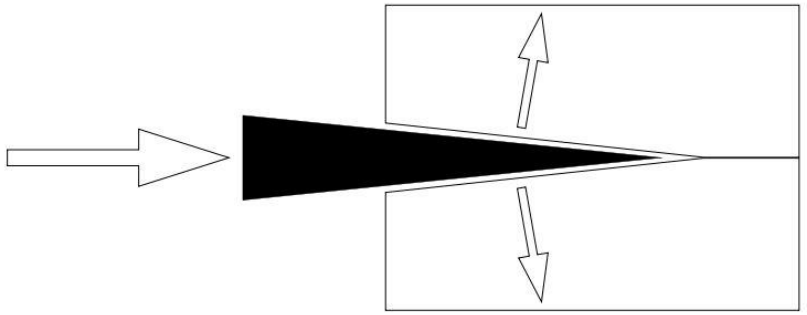
年金機構の情
教訓を生かす
IT部門
危機対応
開催日時
2016年7月13日
開始場所
ベルサール神田
SAS Institute Japa
セルフサービス ×
=マネージド
レッドハット

©2016 EXA CORPORATION

まとめ

WEXをコグニティブコンピューティングへの『楔（くさび）』として使う（のでは）

- まず検索、次に分析、使い方が見つかればSI
- Watson Explorerをコグニティブなシステムの基盤として使う
  - コーパスの設計支援に
  - システムのUIとして
    - Watson Developer Cloud と連携させる
- 過去に成功したユースケースの再販
  - カスタム開発中心
  - 個別案件ごとにカスタム化したWatson



©2016 EXA CORPORATION

おわり

I'm lovin' it

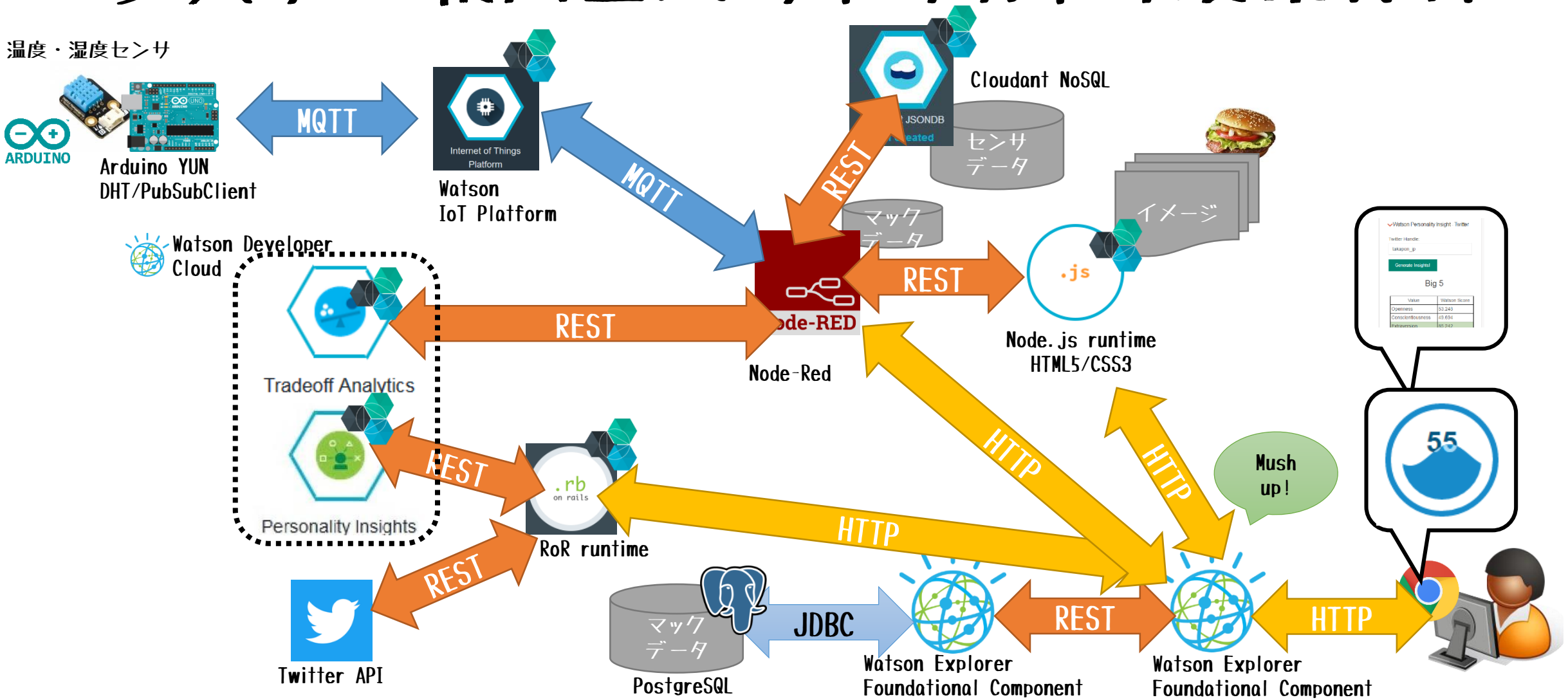
本資料に記載の企業名称、製品名称は、米国およびその他の国における各社の登録商標です。

©2016 EXA CORPORATION

おまけ

【参考】360° ビュー

システム構成図:マクドナルド栄養素分析



©2016 EXA CORPORATION

参考資料

- 「Watsonとの対話」渡辺謙＋IBM Watson CM動画リンク
  - 現在も放送中のWatson CMのYoutube動画
- NHKサイエンスZERO「人工知能の大革命!ディープラーニング」動画リンク
  - ディープラーニングを取り上げたNHKの番組、ただし有料
- Watson Summit2016 A3-1 セッション動画リンク
  - IBM担当者によるWEX概要紹介動画、資料ファイルを参照したい場合は要ログイン
- Creating and Configuring Watson Explorer Application Builder Applications Knowledge Centerリンク
  - Application Builderをインストールして使う方法が記述、ただし英語
- IBMのWatsonが実戦デビュー、三井住友銀行とみずほ銀行で年明けにも 記事リンク
  - ITProの記事リンクです、無料です

©2016 EXA CORPORATION