

優れた検索アプリケーションを作る

マークロジック株式会社
シニアソリューションアーキテクト
伊藤 滋伸

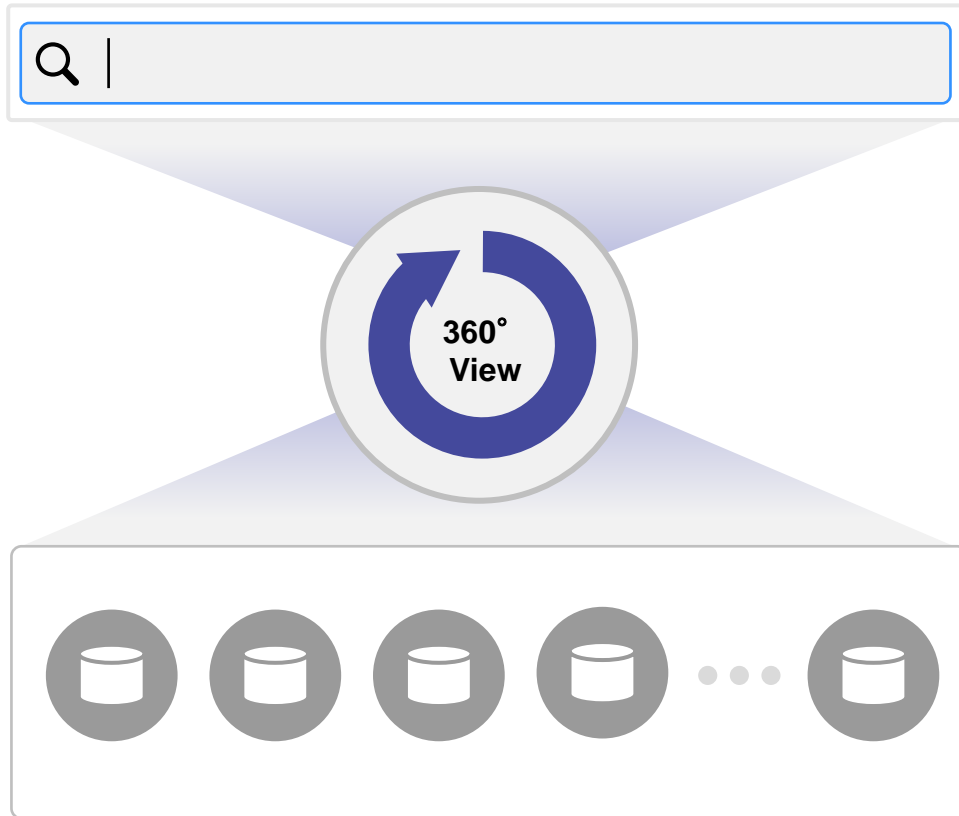


アジェンダ

- エンタープライズ検索
- 検索に求められるアバウトネス (Aboutness)
- 検索アプリケーションへのアバウトネスの実装例
 - セマンティック検索
 - データ統合検索
- まとめ

今日の問題

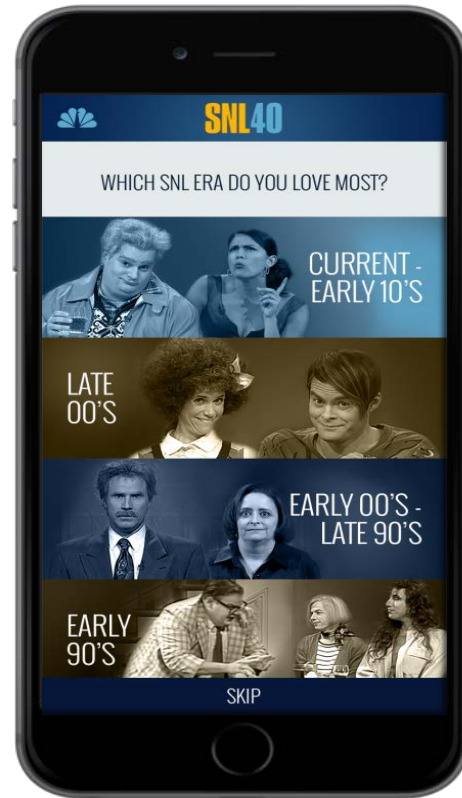
エンタープライズ検索は web検索とは違う



MarkLogicが提供するもの

エンタープライズ用の 素晴らしい検索

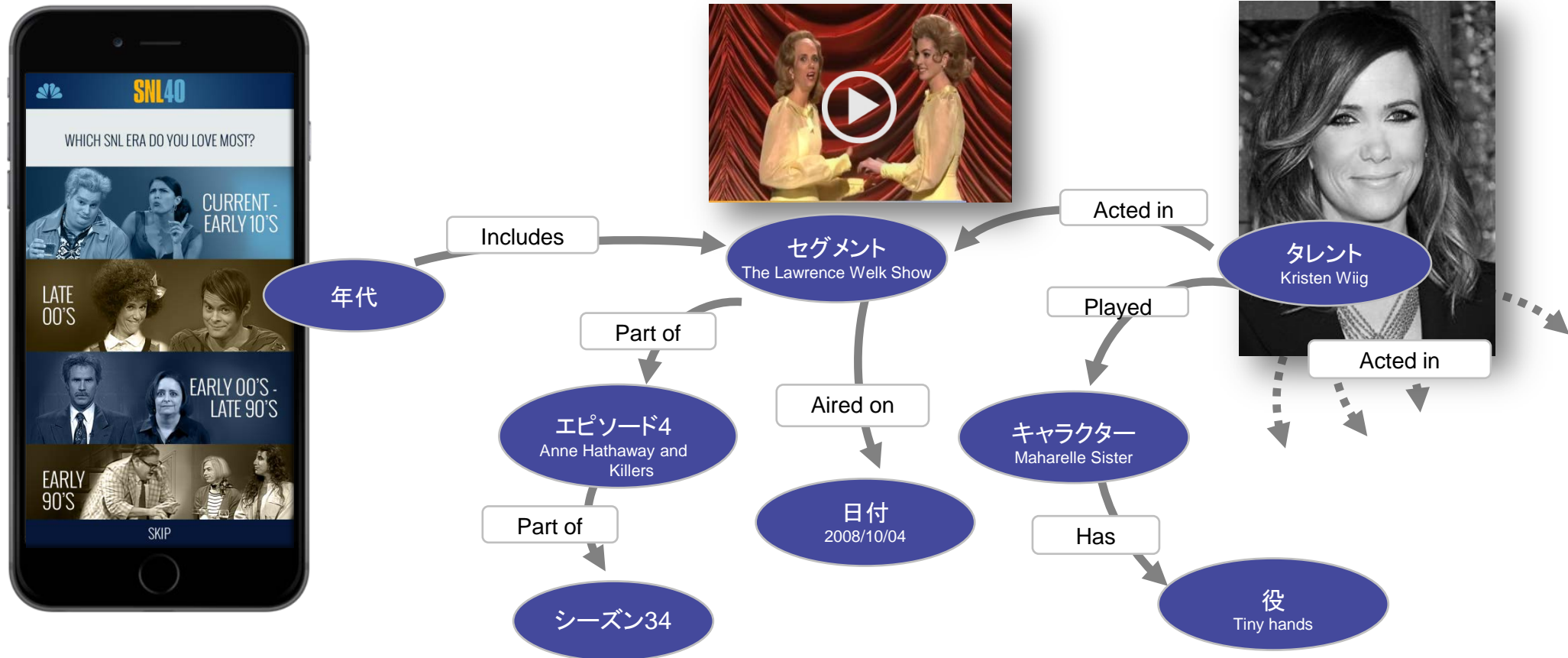
NBCユニバーサル – SNL40アプリ



- スマートコンテンツ
 - コンテンツのエンリッチ化と対象化
 - リコメンデーションを常にチューニング

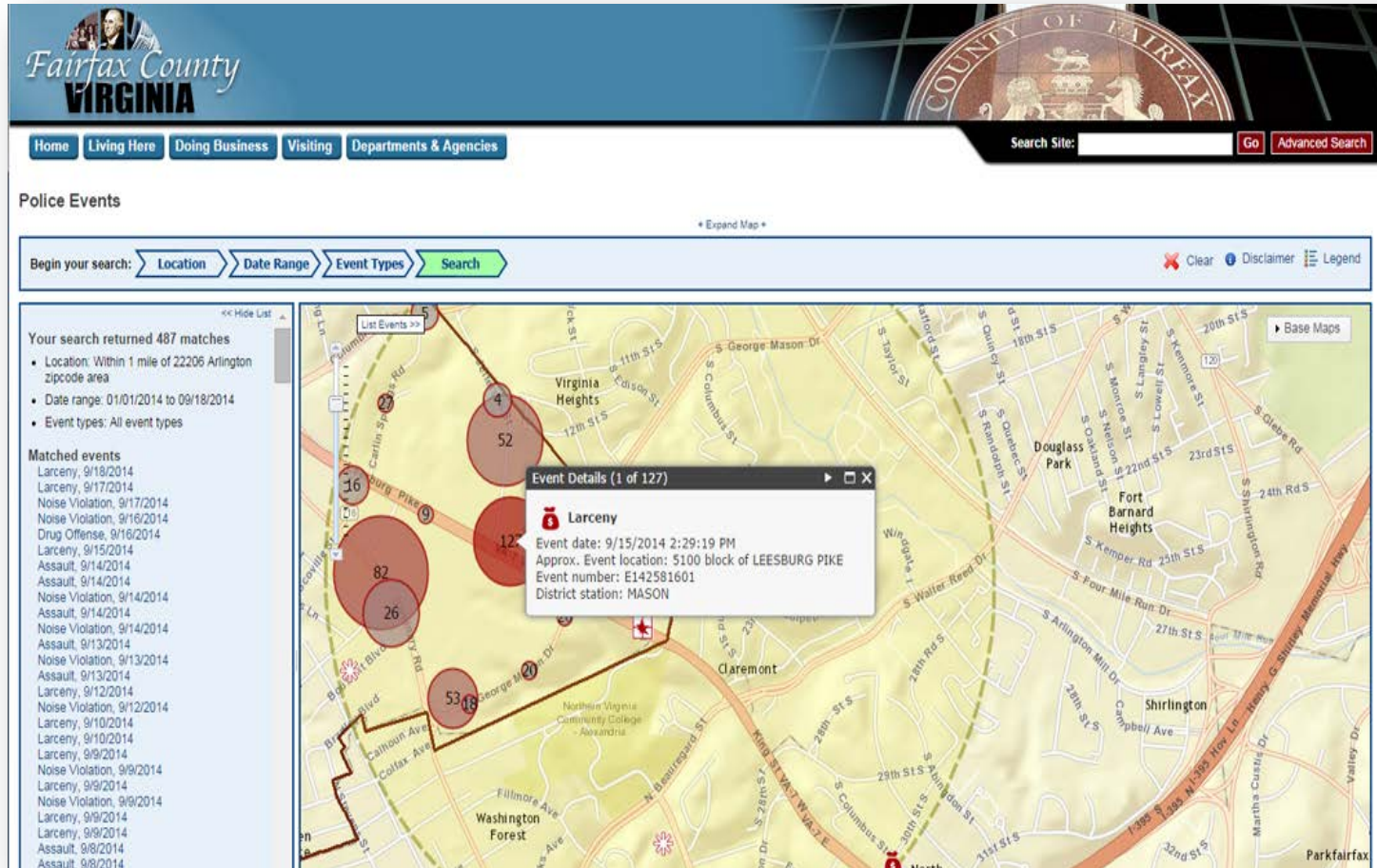
- 5550個以上のスケッチ(動画)

NBCユニバーサル – SNL40アプリ



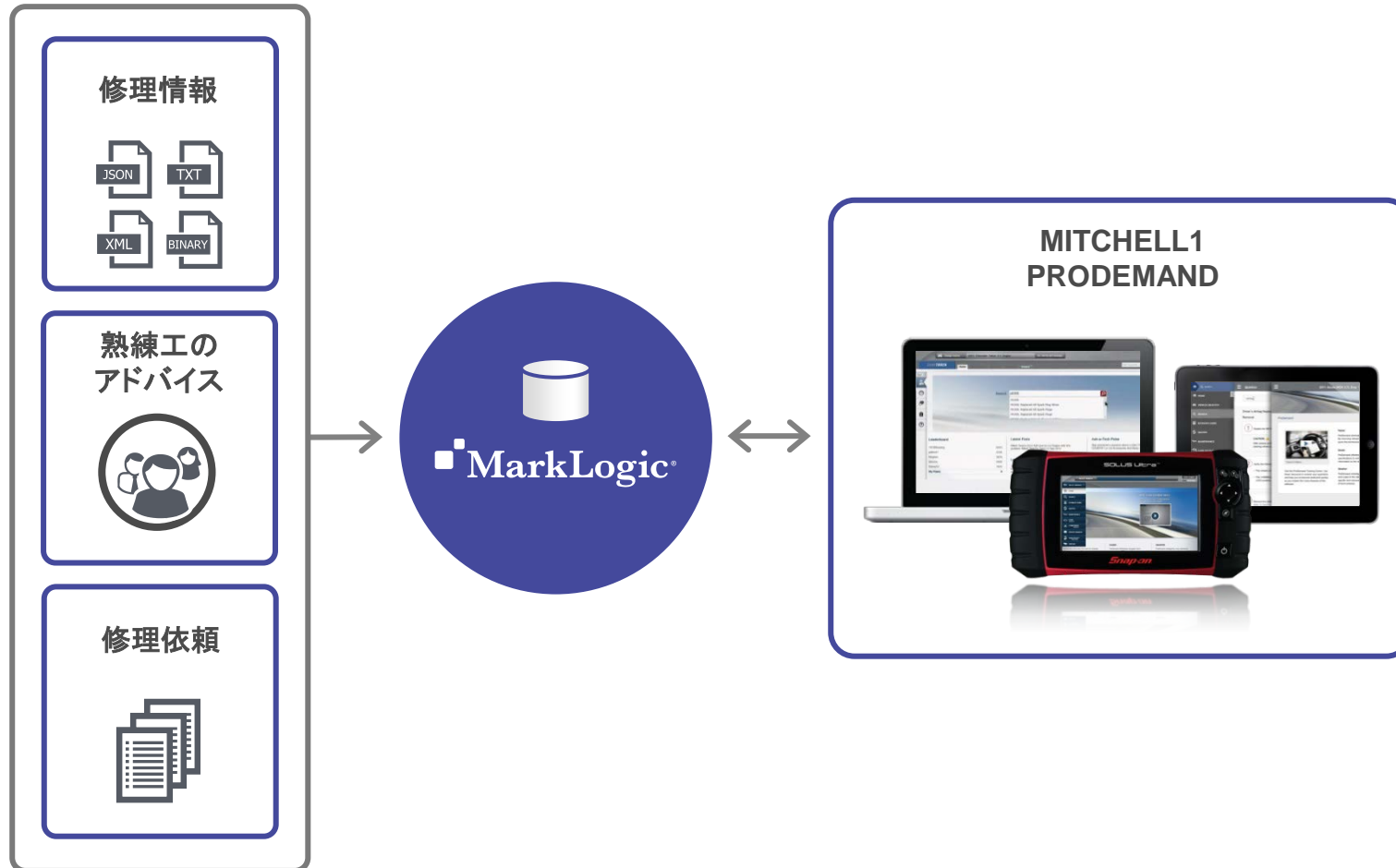
スマートコンテンツ: コンテンツのエンリッチ化と対象化
 リコメンデーションを常にチューニング

米国フェアファックス郡 – 犯罪データのマッピング



- 多様なデータ
 - 50万+ GIS位置データポイント
 - 通報
- 多様なソース
 - 800+ GIS位置レイヤー
 - 構造化 & 非構造化

Mitchell1 – 自動車修理情報を一元管理



- 構造化 & 非構造化データ
- 洗練された検索
- マルチデバイスへの提供

素晴らしい検索アプリケーションはどう作られるか？



データ

- 構造化 & 非構造化
- マルチモデルアプローチ



クエリ

- 重要度
- スニペット
- 強調表示
- オートコンプリート
- スペルの自動修正
- 同義語
- 分析、ドリルダウン



プラットフォーム

- アラート
- トランザクション
- セキュリティ
- リアルタイム
- 言語
- 設定可能
- 開発プラットフォーム

◆◆ コンテキスト

◆◆ プレゼンテーション

今日の問題

アバウトネス (Aboutness)

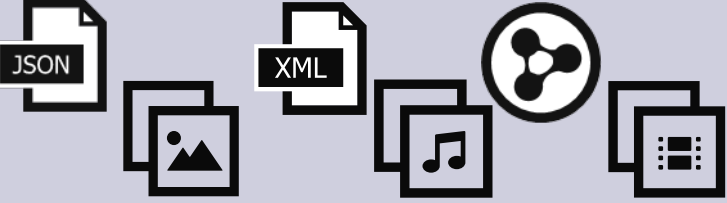
アバウトネス – 検索成功の鍵

- このドキュメントは何に関するものか？
- このサブミットされたクエリは何に関するものか？
- このユーザーは何をしようとしているのか？
- このレスポンスは何に関するものか？

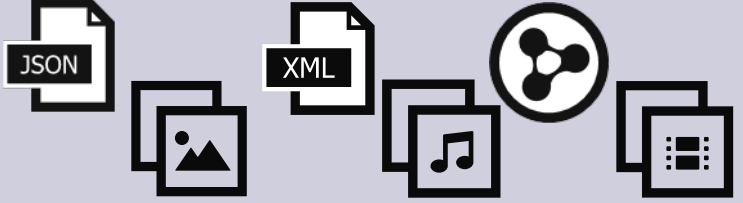
MarkLogicにおけるアバウトネスの実装

データ	このドキュメントは何に関するものか？

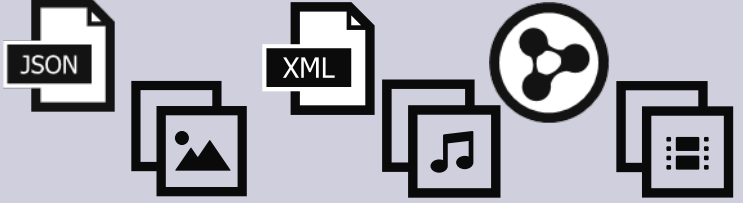

MarkLogicにおけるアバウトネスの実装

データ	

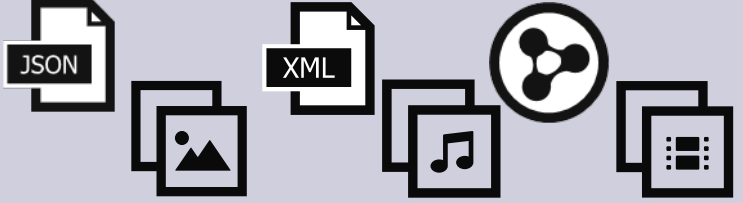

MarkLogicにおけるアバウトネスの実装

データ	
クエリ	このクエリは何に関するものか？

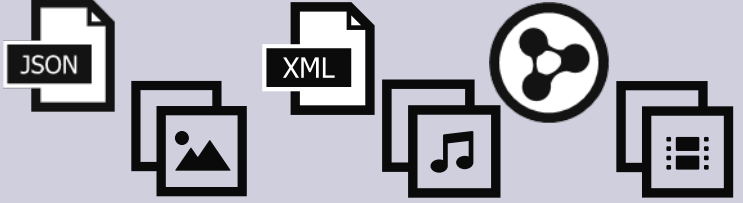


MarkLogicにおけるアバウトネスの実装

データ	
クエリ	

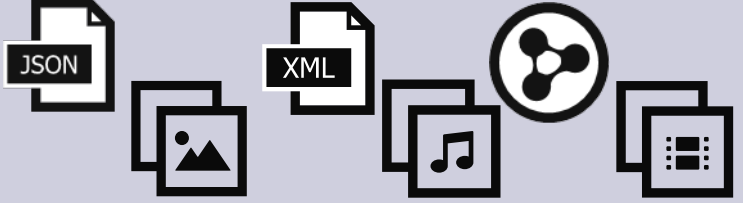


MarkLogicにおけるアバウトネスの実装

データ	
クエリ	
コンテキスト	このユーザーは何をしようとしているのか？

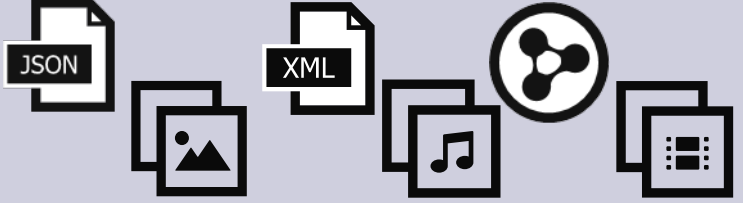



MarkLogicにおけるアバウトネスの実装

データ	
クエリ	
コンテキスト	

MarkLogicにおけるアバウトネスの実装

データ	
クエリ	
コンテキスト	
結果	このレスポンスは何に関するものか？

MarkLogicにおけるアバウトネスの実装

データ	
クエリ	
コンテキスト	
結果	

素晴らしい検索アプリケーションへのアバウトネスの実装

- データ
 - カテゴリ化、エンベロープ構造
- クエリ
 - ダイアログとしてのクエリ
- コンテキスト
 - 情報リソースの構築
- 結果
 - 的確なレスポンス

素晴らしい検索アプリケーションへのアバウトネスの実装

- データ
 - カテゴリ化、エンベロープ構造
- クエリ
 - ダイアログとしてのクエリ
- コンテキスト
 - 情報リソースの構築
- 結果
 - 的確なレスポンス

業界知識 – アバウトネス作成のためのツール

- タクソノミー - 組織内の語彙
- オントロジー - ドキュメントを関連するデータやドキュメントと結びつける
- シソーラス
- 業務部門、人、製品
- 一般公開されている語彙 - Geographic、DBPedia

タクソミー

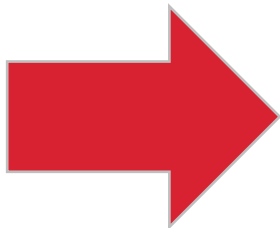
ドキュメント内で探したいターム

```
<rdf:Description rdf:about="http://vocabulary.worldbank.org/taxonomy/333">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <dcterms:creator>wb</dcterms:creator>
  <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-05-
17T15:05:21Z</dcterms:created>
  <skos:prefLabel xml:lang="en">Non-Government Bond Markets</skos:prefLabel>
  <dcterms:contributor>wb</dcterms:contributor>
  <skos:broader rdf:resource="http://vocabulary.worldbank.org/taxonomy/1918"/>
  <skos:altLabel xml:lang="en">Nongovernment Bonds</skos:altLabel>
  <skos:altLabel xml:lang="en">Non-government Bonds</skos:altLabel>
  <skos:altLabel xml:lang="en">Nongovernment Bond Markets</skos:altLabel>
  <skos:altLabel xml:lang="en">Corporate Bond Markets</skos:altLabel>
  <dcterms:hasVersion rdf:datatype="http://www.w3.org/2001/XMLSchema#int">10</dcterms:hasVersion>
  <skos:altLabel xml:lang="en">Private Bond Markets </skos:altLabel>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2014-12-
22T01:47:25Z</dcterms:modified>
</rdf:Description>
```

カテゴリ化の実装

■ タクソノミーエンティティをクエリ化

```
<rdf:Description rdf:about="http://vocabulary.worldbank.org/taxonomy/333">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <dcterms:creator>wb</dcterms:creator>
  <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2013-05-
17T15:05:21Z</dcterms:created>
  <skos:prefLabel xml:lang="en">Non-Government Bond Markets</skos:prefLabel>
  <dcterms:contributor>wb</dcterms:contributor>
  <skos:broader rdf:resource="http://vocabulary.worldbank.org/taxonomy/1918"/>
  <skos:altLabel xml:lang="en">Nongovernment Bonds</skos:altLabel>
  <skos:altLabel xml:lang="en">Non-government Bonds</skos:altLabel>
  <skos:altLabel xml:lang="en">Nongovernment Bond Markets</skos:altLabel>
  <skos:altLabel xml:lang="en">Corporate Bond Markets</skos:altLabel>
  <dcterms:hasVersion rdf:datatype="http://www.w3.org/2001/XMLSchema#int">10</dcterms:hasVersion>
  <skos:altLabel xml:lang="en">Private Bond Markets </skos:altLabel>
  <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2014-12-
22T01:47:25Z</dcterms:modified>
</rdf:Description>
```



```
<cts:or-query xmlns:cts="http://marklogic.com/cts">
  <cts:near-query distance="5">
    <cts:near-query distance="5">
      <cts:element-word-query weight="10">
        <cts:element
          xmlns:_1="http://marklogic.com/solutions/obi/source">_1
          :title</cts:element>
          <cts:text xml:lang="en">Non-Government</cts:text>
          <cts:option>case-insensitive</cts:option>
          <cts:option>punctuation-insensitive</cts:option>
          <cts:option>stemmed</cts:option>
        </cts:element-word-query>
      <cts:element-word-query weight="10">
        <cts:element
          xmlns:_1="http://marklogic.com/solutions/obi/source">_1
          :title</cts:element>
          <cts:text xml:lang="en">Bond</cts:text>
          <cts:option>case-insensitive</cts:option>
          <cts:option>punctuation-insensitive</cts:option>
          <cts:option>stemmed</cts:option>
        </cts:element-word-query>
      </cts:near-query>
```


セマンティックの世界

リンクオープンデータ
(無料で公開されているセマンティックのファクト)



プロプライエタリなセマンティックのファクト
(組織ごとのファクトとタクソミー)

CLASS OF DRUG	ADVERSE DRUG REACTIONS OR EFFECTS
1) MORPHINE	Dyspnea, constipation, dryness of mouth, mental clouding, headache and fatigue. Intolerance and respiratory depression.
1) PETHIDINE and derivatives	Swelling, euphoria, dizziness, dry mouth, weakness and pallidation. Anaphylactic shock leading to respiratory collapse and bronchospasm, coma or tetanus.
2) HYPNOTIC SEDATIVES	Headache, nausea, vomiting, itchy skin lesions like urticaria, slurred speech, ataxia, impaired reflexes, anxiety, restlessness, tremors, abdominal cramps, CNS depression, particularly respiratory and oculomotor collapse, rapid pulse.
3) ANTIEMETIC ANAESTHETICS	Skin rashes, urticaria, angioneurotic edema. GI problems such as dyspepsia, nausea, vomiting and sometimes GI bleeding. Reduction in plasma prothrombin level and interference with prostaglandin synthesis.
4) NON-STEROIDAL ANTI-INFLAMMATORY DRUGS (NSAIDs)	Skin reactions, anemia, hemolytic, large doses cause liver damage.
5) PAINKILLERS	Peptic ulcer, febrile reactions such as anemia, agranulocytosis, edema, displacement of erythrocytes due to high protein binding, leukemia.
6) OTHER NSAIDs like indometacin, ibuprofen, naproxen.	Headache, mental confusion, blurring of vision, depression, peptic ulcer, liver damage, renal damage.

セマンティックの世界



ドキュメント内のファクト
(メタデータの一部やオーサリングツールで追加されたもの)

```
Documents can contain triples
<article>
<meta>
<title>Man bites dog</title>
<sem:triple>
<sem:subject>http://example.org/news/42</sem:subject>
<sem:predicate>http://example.org/published</sem:predicate>
<sem:object>2013-09-10</sem:object>
</sem:triple>
...
```

ドキュメントの世界

一般的なテキストからのファクト
(セマンティックエンリッチメントにより生成)



エンベロープ – Know Your Data

- 読み込まれたさまざまなタイプのドキュメントを標準化
 - 一般的な名称が異なる可能性 (例: Title、Heading、Subject)
 - 検索対象の要素を一般的な形式に正規化
 - 同一ドキュメントの複数バージョン (例: 翻訳)
- メタデータに対するインデックス検索 (ファセット)
- コンテンツに対する全文検索



素晴らしい検索アプリケーションへのアバウトネスの実装

- データ
 - カテゴリ化、エンベロープ構造
- クエリ
 - ダイアログとしてのクエリ
- コンテキスト
 - 情報リソースの構築
- 結果
 - 的確なレスポンス

ダイアログを通じたアバウトネスの作成

このクエリは何に関するものか？

- 入力支援
 - レキシコンに対するマッチ
 - 検索履歴に対するマッチ(ユーザー、グローバル)
- 背景を考えた検索
 - フレーズの特定
 - オブジェクトマッチング
 - クエリのブースト

ダイアログを通じたアバウトネスの作成(続き)

このクエリは何に関するものか？

- 修正と確定
 - スペルチェック
 - シソーラスによる拡張
 - セマンティックによる拡張
- ドリルダウン
 - フォローアップの検索ボックス
 - ファセット
 - 即時の制約

検索アプリケーションへのアバウトネスの実装例

- セマンティックを使った検索
 - 目的 (Object)
 - 同義、同じクラス、関連性
 - 推論

- データ統合の優れたアプリケーション
 - レンジインデックス
 - フィールド
 - 分析 (グラフ)

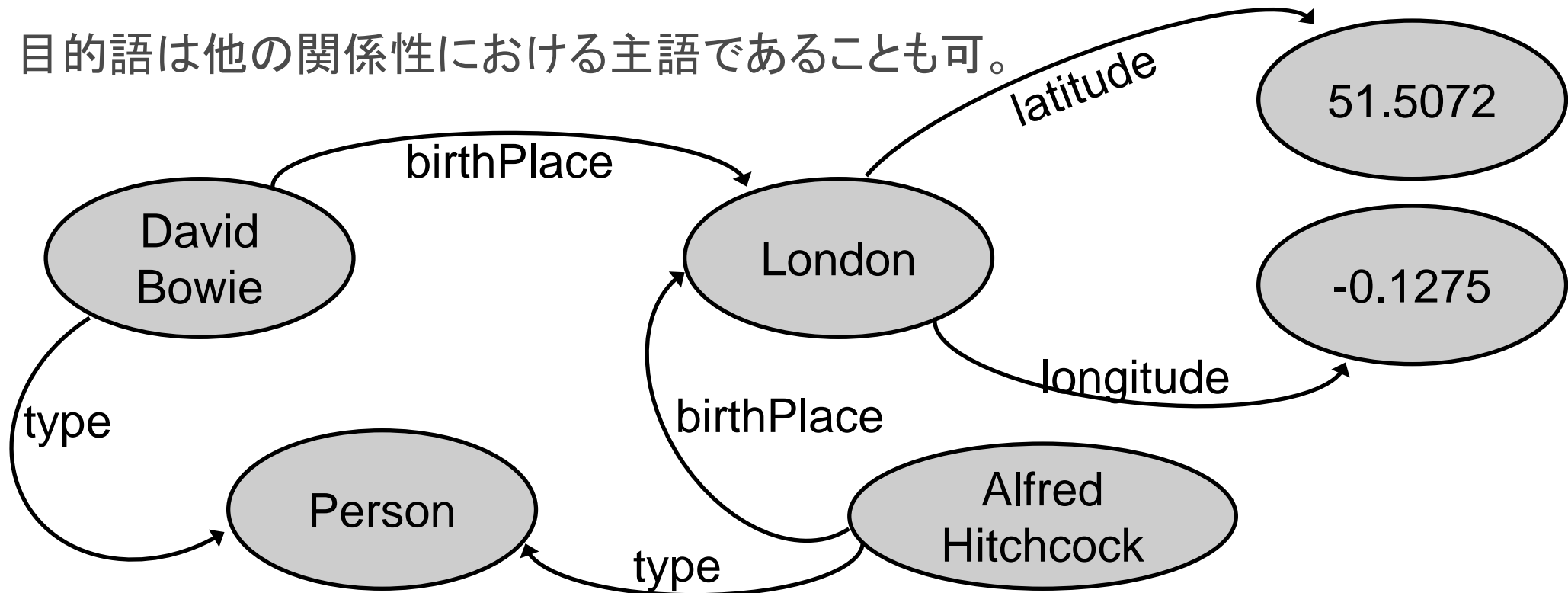
今日の問題

セマンティック検索

RDFグラフ

- グラフ

- RDFトリプルのステートメントやパターンの集まり。グラフベースのRDFモデルでは、ノードは主語/目的語のリソースを表し、ノードの繋がりが述語。
- 目的語は他の関係性における主語であることも可。



推論 (inference)

- 一連の前提から論理的な結論を導く
- データとオントロジーに基づき新しい関係性を発見
- 例:
 - これらの主語のうちミュージシャンはどれか？
 - 推論なしだと結果はゼロ
 - 推論を使うと、これらの主語全員がミュージシャンだとわかる

トリプルの例 :

David Bowie	type	Singer
Eric Clapton	type	Guitarist
Beethoven	type	Composer

オントロジーの例 :

Composer	equivalentClass	Musician
Singer	subClassOf	Musician
Guitarist	subClassOf	Musician

具体的なアプリケーション例① ～エボラ出血熱における情報を調べる～

- エボラ出血熱に関して情報を調べる
 - 普通に検索する(Content) vs 意味を検索する(Object)
 - 関連した情報(誰が?いつ?何を?どこで?)をまとめてオブジェクトを検索(意味検索)
- *Kent Brantly*
 - Kent Brantlyとは誰か?
 - 彼はどこにいたか?
 - 彼は誰と関連したか?
 - 彼はなぜエボラ出血熱にかかったか?
 - 彼は何をしたか?
- グラフそして(関連)ドキュメント
 - セマンティックから関連性を引っ張り出す
 - 関連しているドキュメントをまとめて引っ張り出す

具体的なアプリケーション例② ～糖尿病の臨床データを探る～

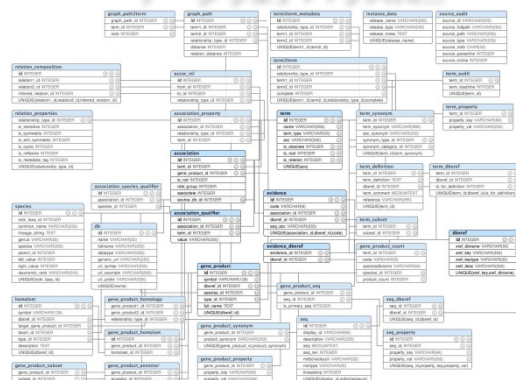
- 最新の医療ニュースを調べる
 - 新しい糖尿病治療薬
 - 試験データ(同じ薬で効用の高い臨床データ)
- 自身の医療機関において糖尿病治療薬データを調べる
 - Twitterや全ての医療データ
 - 患者データ
 - 製品での絞り込み
- 患者データ
 - 血圧、貧血(糖尿病のインスリン薬には針がある)
 - 投薬データ、分析

このような知識は組織内のどこに存在しているのか？

ある部分は頭の中に...



ある部分はRDBMSや
データウェアハウスに...



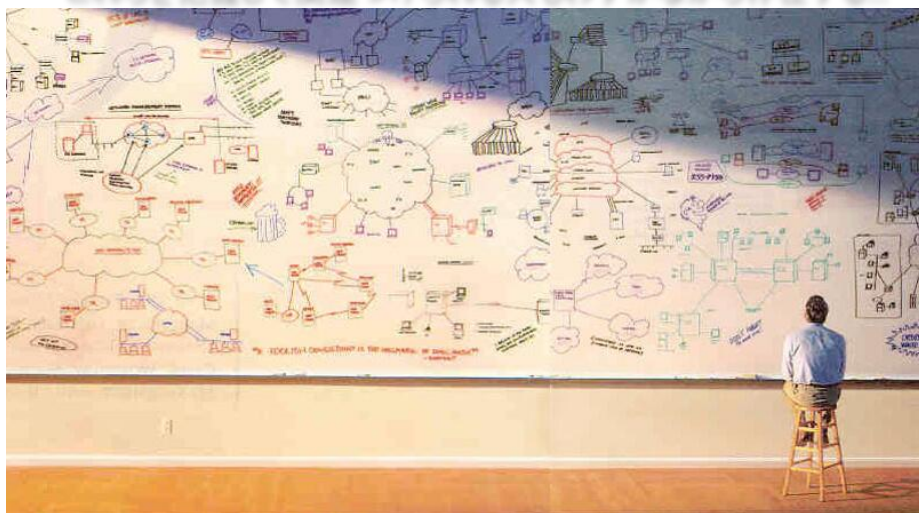
ある部分はナレッジマネジメント
システムに...



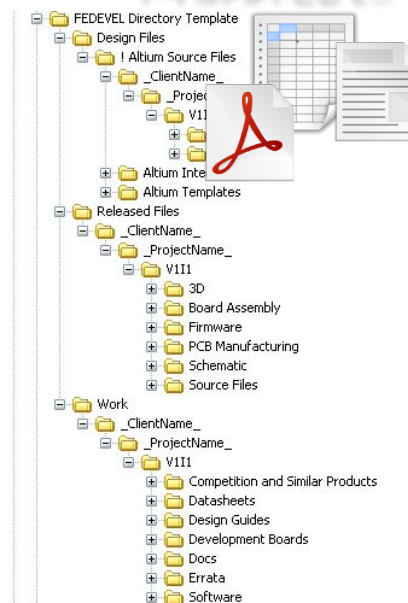
ある部分は永遠に失われる...



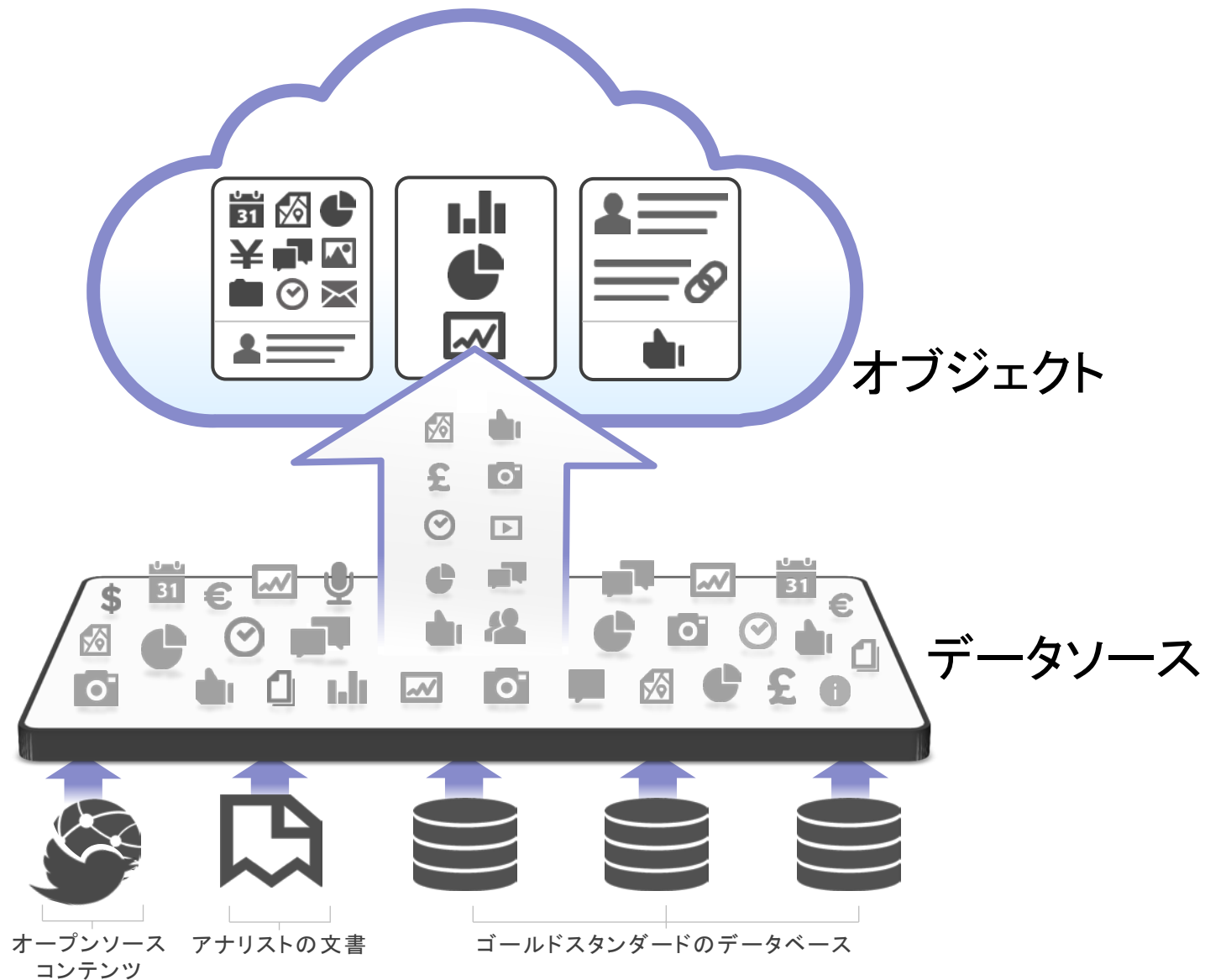
これらを把握して関連づけられる人もわずかながら存在する...



ある部分は共有ドライブの
ドキュメントとして...



論理的アーキテクチャ



今日の問題

データ統合検索

Purchase Order

P.O. Number
P.O. Date

Customer

Name
Address1
Address2
Suburb
Post Code

Delivery

Name
Address1
Address2
Suburb
Post Code

Line #	Part #	Description	Qty.	Item Cost
1	A56-2	Variable Speed 200kw Motor	1	\$3,502.43
2	C12-2	Transformer – 240v	2	\$564.45
3	C738	Fitting Bracket – C12-2	2	\$12.35

Purchase Order

P.O. Number

CY3456

P.O. Date

23/05/2015

Customer

Name Albert Johns

Address1 52 Horizon Drive

Address2

Suburb Botany

Post Code 2157

Delivery

Name Ken Wilson

Address1 Unit 4

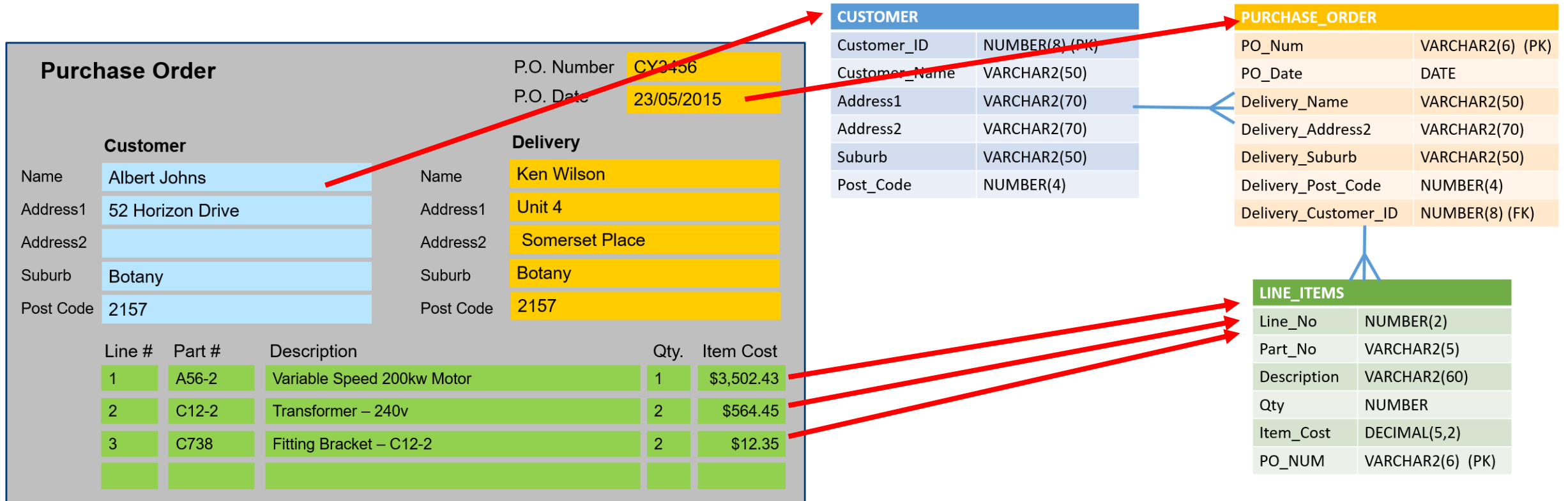
Address2 Somerset Place

Suburb Botany

Post Code 2157

Line #	Part #	Description	Qty.	Item Cost
1	A56-2	Variable Speed 200kw Motor	1	\$3,502.43
2	C12-2	Transformer – 240v	2	\$564.45
3	C738	Fitting Bracket – C12-2	2	\$12.35

リレーショナルデータモデルによる設計



ドキュメントモデルによる設計

Purchase Order				
		P.O. Number	CY3456	
		P.O. Date	23/05/2015	
Customer			Delivery	
Name	Albert Johns		Name	Ken Wilson
Address1	52 Horizon Drive		Address1	Unit 4
Address2			Address2	Somerset Place
Suburb	Botany		Suburb	Botany
Post Code	2157		Post Code	2157
Line #	Part #	Description	Qty.	Item Cost
1	A56-2	Variable Speed 200kw Motor	1	\$3,502.43
2	C12-2	Transformer – 240v	2	\$564.45
3	C738	Fitting Bracket – C12-2	2	\$12.35

```

<?xml version="1.0"?>
<PurchaseOrderNumber>="CY3456"</PurchaseOrderNumber>
<OrderDate>"2015-05-23"</OrderDate>
<CustomerName>Albert Johns</CustomerName>
<Address Type="Customer">
  <Street>52 Horizon Drive</Street>
  <Suburb>Botany</Suburb>
  <State>NSW</State>
  <PostCode>2157</PostCode>
</Address>
<LinItems>
  <LinItem>1</LinItem>
  <Item PartNumber="A56-2">
    <ProductDesc>Variable Speed 200km Motor</ProductDesc>
    <Quantity>1</Quantity>
    <ItemCost>3502.43</ItemCost>
  </LinItem>
  <LinItem>2</LinItem>
  <Item PartNumber=C12-2>
    <ProductDesc>Transformer – 240v</ProductDesc>
    <Quantity>2</Quantity>
    <ItemCost>564.45</ItemCost>
  </LinItem>
  <LinItem>3</LinItem>
  <Item PartNumber=C738>
    <ProductDesc>Fitting Bracket – C12-2</ProductDesc>
    <Quantity>2</Quantity>
    <ItemCost>12.35</ItemCost>
  </LinItem>
</PurchaseOrder>
  
```

リレーショナルからドキュメントへ

CUSTOMER	
Customer_ID	NUMBER(8) (PK)
Customer_Name	VARCHAR2(50)
Address1	VARCHAR2(70)
Address2	VARCHAR2(70)
Suburb	VARCHAR2(50)
Post_Code	NUMBER(4)

PURCHASE_ORDER	
PO_Num	VARCHAR2(6) (PK)
PO_Date	DATE
Delivery_Name	VARCHAR2(50)
Delivery_Address2	VARCHAR2(70)
Delivery_Suburb	VARCHAR2(50)
Delivery_Post_Code	NUMBER(4)
Delivery_Customer_ID	NUMBER(8) (FK)

LINE_ITEMS	
Line_No	NUMBER(2)
Part_No	VARCHAR2(5)
Description	VARCHAR2(60)
Qty	NUMBER
Item_Cost	DECIMAL(5,2)
PO_NUM	VARCHAR2(6) (PK)

```

<?xml version="1.0"?>
<PurchaseOrderNumber>="CY3456"</PurchaseOrderNumber>
<OrderDate>"2015-05-23"</OrderDate>
<CustomerName>Albert Johns</CustomerName>
<Address Type="Customer">
  <Street>52 Horizon Driver</Street>
  <Suburb>Botany</Suburb>
  <State>NSW</State>
  <PostCode>2157</PostCode>
</Address>
<LineItems>
  <LineItem>1</LineItem>
  <Item PartNumber="A56-2">
    <ProductDesc>Variable Speed 200km Motor</ProductDesc>
    <Quantity>1</Quantity>
    <ItemCost>3502.43</ItemCost>
  </LineItem>
  <LineItem>2</LineItem>
  <Item PartNumber=C12-2>
    <ProductDesc>Transformer – 240v</ProductDesc>
    <Quantity>2</Quantity>
    <ItemCost>564.45</ItemCost>
  </LineItem>
  <LineItem>3</LineItem>
  <Item PartNumber=C738>
    <ProductDesc>Fitting Bracket – C12-2</ProductDesc>
    <Quantity>2</Quantity>
    <ItemCost>12.35</ItemCost>
  </LineItem>
</PurchaseOrder>
  
```

- 最適なデータモデル
- “行”をドキュメントへ
- ビジネスの“モデル”で再構成

“行”のドキュメント化

```
<?xml version="1.0"?>
<PurchaseOrderNumber>="CY3456"</PurchaseOrderNumber>
<OrderDate>"2015-05-23"</OrderDate>
```

```
<?xml version="1.0"?>
<CustomerName>Albert Johns</CustomerName>
<Address Type="Customer">
  <Street>52 Horizon Driveri</Street>
  <Suburb>Botany</Suburb>
  <State>NSW</State>
  <PostCode>2157</PostCode>
</Address>
```

- 元データを各々のドキュメントとして管理
- “データ”を扱う
- 各々の行をエンリッチ(セキュリティ、リンク、データの意味、、、)

```
<?xml version="1.0"?>
<LineItem>
  <Item PartNumber=C12-2>
    <ProductDesc>Transformer – 240v</ProductDesc>
    <Quantity>2</Quantity>
    <ItemCost>564.45</ItemCost>
  </LineItem>
```

```
<?xml version="1.0"?>
<LineItem>3</LineItem>
<Item PartNumber=C738>
  <ProductDesc>Fitting Bracket – C12-2</ProductDesc>
  <Quantity>2</Quantity>
  <ItemCost>12.35</ItemCost>
</LineItem>
```

```
<?xml version="1.0"?>
<LineItem>3</LineItem>
<Item PartNumber=C738>
  <ProductDesc>Fitting Bracket – C12-2</ProductDesc>
  <Quantity>2</Quantity>
  <ItemCost>12.35</ItemCost>
</LineItem>
```

“行”のドキュメント化

- ビジネスモデルを反映させる

```
<?xml version="1.0"?>
<PurchaseOrderNumber>="CY3456"</PurchaseOrderNumber>
<OrderDate>"2015-05-23"</OrderDate>
```

```
<CustomerName>Albert Johns</CustomerName>
<Address Type="Customer">
  <Street>52 Horizon Driveri</Street>
  <Suburb>Botany</Suburb>
  <State>NSW</State>
  <PostCode>2157</PostCode>
</Address>
```

- ビジネスオブジェクトを反映
- “行”データから、あるまとまったビジネスレイヤを表すように作成

```
<LineItem>
  <LineItem>
    <Item PartNumber=C12-2>
      <ProductDesc>Transformer – 240v</ProductDesc>
      <Quantity>2</Quantity>
      <ItemCost>564.45</ItemCost>
    </LineItem>
  </LineItem>
  <LineItem>3</LineItem>
  <Item PartNumber=C738>
    <ProductDesc>Fitting Bracket – C12-2</ProductDesc>
    <Quantity>2</Quantity>
    <ItemCost>12.35</ItemCost>
  </LineItem>
</PurchaseOrder>
```

“エンベロープ”によるドキュメント化

```
<?xml version="1.0"?>
<PurchaseOrderNumber>="CY3456"</PurchaseOrderNumber>
<OrderDate>"2015-05-23"</OrderDate>
```

```
<?xml version="1.0"?>
<CustomerName>Albert Johns</CustomerName>
<Address Type="Customer">
  <Street>52 Horizon Driveri</Street>
  <Suburb>Botany</Suburb>
  <State>NSW</State>
  <PostCode>2157</PostCode>
</Address>
```

- 重複データを除いてビジネスオブジェクトでまとめる
- ドキュメントはエンベロープ(封筒)の考え方で整理
- ドキュメントレベルではなくエンベロープレベルでエンリッチ

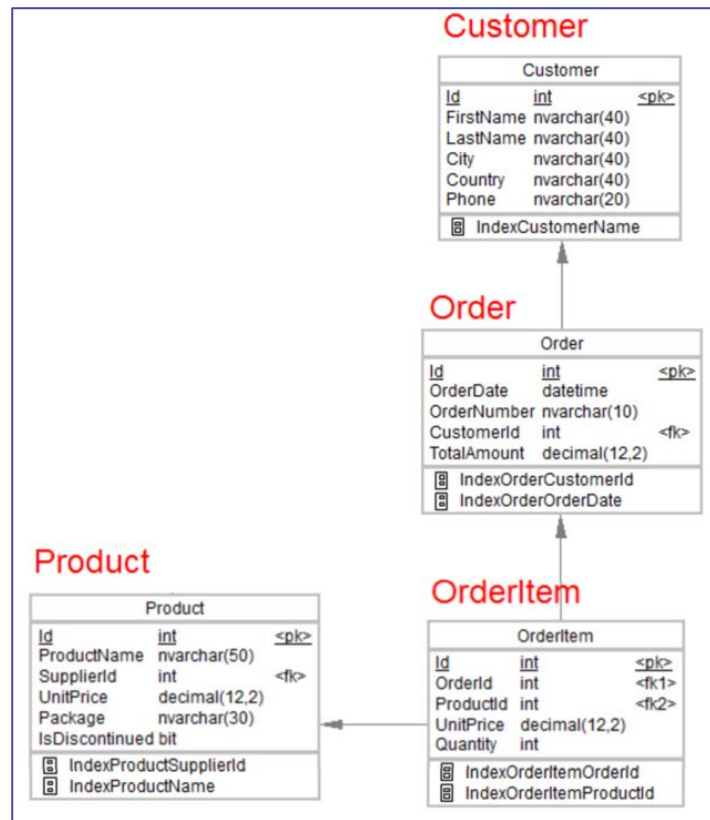
```
<?xml version="1.0"?>
<LineItem>
  <LineItem>
    <Item PartNumber=C12-2>
      <ProductDesc>Transformer – 240v</ProductDesc>
      <Quantity>2</Quantity>
      <ItemCost>564.45</ItemCost>
    </LineItem>
  </LineItem>
  <LineItem>3</LineItem>
  <Item PartNumber=C738>
    <ProductDesc>Fitting Bracket – C12-2</ProductDesc>
    <Quantity>2</Quantity>
    <ItemCost>12.35</ItemCost>
  </LineItem>
</PurchaseOrder>
```

データモデルを分かりやすく！

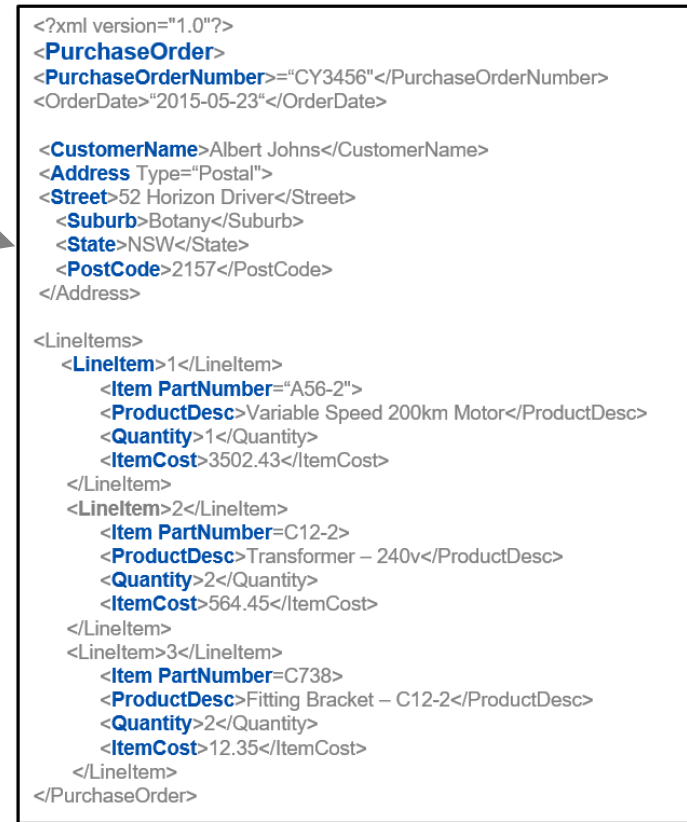
リレーショナルデータモデル

“ビジネスオブジェクト”のドキュメントモデル

“テーブルとスキーマ”



“PurchaseOrder”



search APIのレスポンス



```
<search:response snippet-format="snippet" total="1" start="1" page-length="10" xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns="" xmlns:search="http://marklogi
<search:result index="1" uri="/songs/Eagles+Hotel-California.xml" path="fn:doc(&quot;/songs/Eagles+Hotel-California.xml&quot;)" score="105728" confidence="0.61
  <search:snippet>
    <search:match path="fn:doc(&quot;/songs/Eagles+Hotel-California.xml&quot;)/ts:top-song">
      <search:highlight>Hotel California</search:highlight>
    </search:match>
    <search:highlight>Hotel
California</search:highlight>
  </search:match>
  <search:match path="fn:doc(&quot;/songs/Eagles+Hotel-California.xml&quot;)/ts:top-song/ts:descr/ts:p[1]"><search:highlight>Hotel California</search:highli
of the same name and was...</search:match>
</search:snippet>
  <search:metadata>
    <ts:title href="http://en.wikipedia.org/wiki/Hotel_California_(song)" xmlns:ts="http://marklogic.com/MLU/top-songs">Hotel California</ts:title>
    <ts:artist href="http://en.wikipedia.org/wiki/Eagles" xmlns:ts="http://marklogic.com/MLU/top-songs">Eagles</ts:artist>
    <search:attribute-meta name="last" parent-name="weeks">1977-05-07</search:attribute-meta>
  </search:metadata>
</search:result>
  <search:facet name="artist" type="xs:string">
    <search:facet-value name="Eagles" count="1">Eagles</search:facet-value>
  </search:facet>
  <search:qtext>"Hotel California" AND "The Eagles"</search:qtext>
  <search:metrics>
    <search:query-resolution-time>PT0.011S</search:query-resolution-time>
    <search:facet-resolution-time>PT0.002S</search:facet-resolution-time>
    <search:snippet-resolution-time>PT0.014S</search:snippet-resolution-time>
    <search:metadata-resolution-time>PT0.001S</search:metadata-resolution-time>
    <search:total-time>PT0.029S</search:total-time>
  </search:metrics>
</search:response>
```

レンジインデックス

- タームリストは「Yes / No」型の質問が得意
 - 値→ドキュメントのマッピング
- 以下のような場合は？
 - <price> が\$50以下のドキュメントを探す
 - <date> が1990-01-01から1999-12-31までのドキュメントを探す
- レンジインデックス
 - 値←→ドキュメントのマッピング
 - 値(型あり)。テキスト的なマッチではない
 - 速いレンジクエリ
 - 速いソート
 - 速い値の抽出
 - ファセット化
- レンジインデックスは、MarkLogicの起動時にメモリ上に置かれる

フィールド

- フィールドに名前を付ける
 - 例:「Performer」
 - 含まれる要素:
 - <artist>|<singer>|<group>|<band>
- 特定のインデックス設定を適用

ドキュメント #1

```
<top-song>  
  <artist>  
    The Beatles  
  </artist>  
</top-song>
```

ドキュメント #2

```
<top-song>  
  <singer>  
    Paul Simon  
  </singer>  
</top-song>
```

ドキュメント #3

```
<top-song>  
  <group>  
    Coldplay  
  </group>  
</top-song>
```

ドキュメント #4

```
<top-song>  
  <band>  
    Radiohead  
  </band>  
</top-song>
```


具体的なアプリケーション例③ ～保険会社のデータ統合～

- 保険データを上手く扱う
 - 自動車保険と健康保険
 - データのロード、ファセット検索、分析とドリルダウン
- 他社(恐らく買収した)保険データとの統合
 - フィールド

今日、紹介するアプリケーションはGitHubからダウンロード、設定できます！

<https://github.com/ryanjdew/ml-slush-discovery-app>

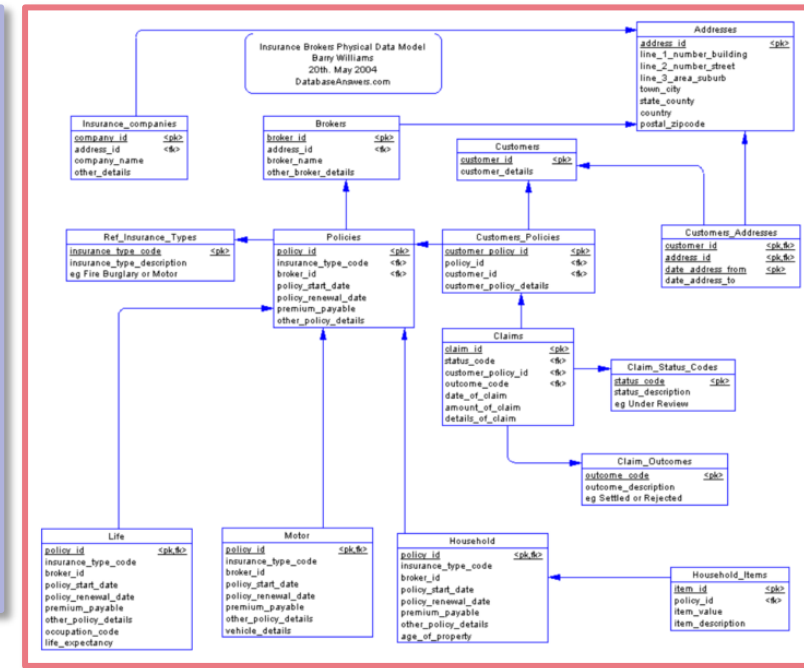
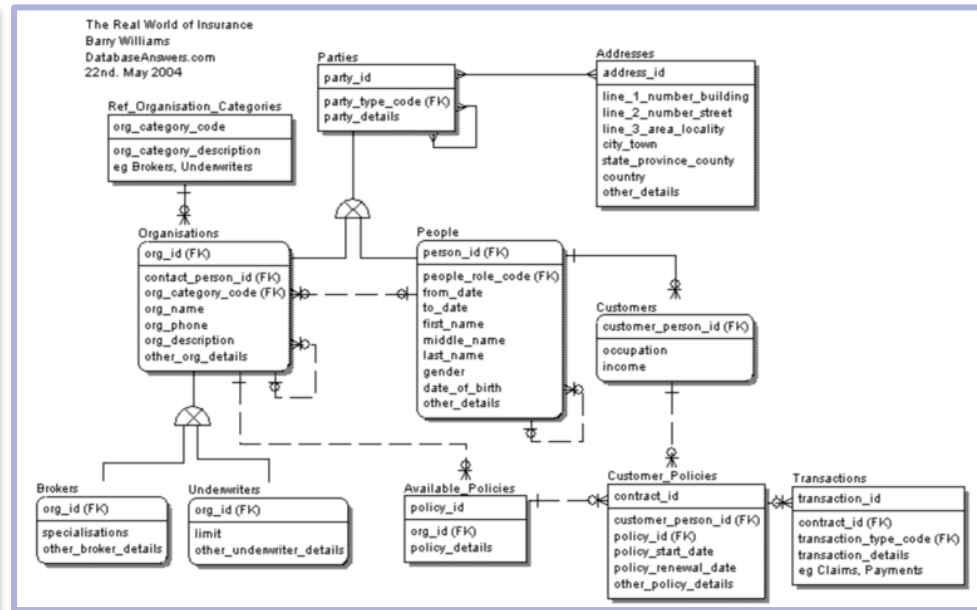
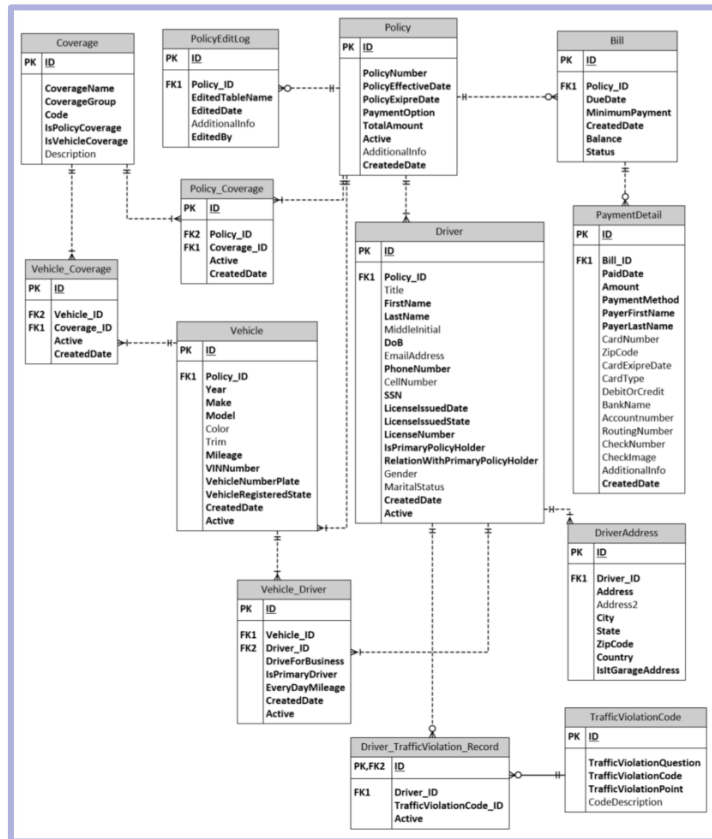
スキーマが違う、ソースが違う、表現が違う

- チャレンジ: データを統合し360° 全方位から検索、データを分析して新しい発見をする

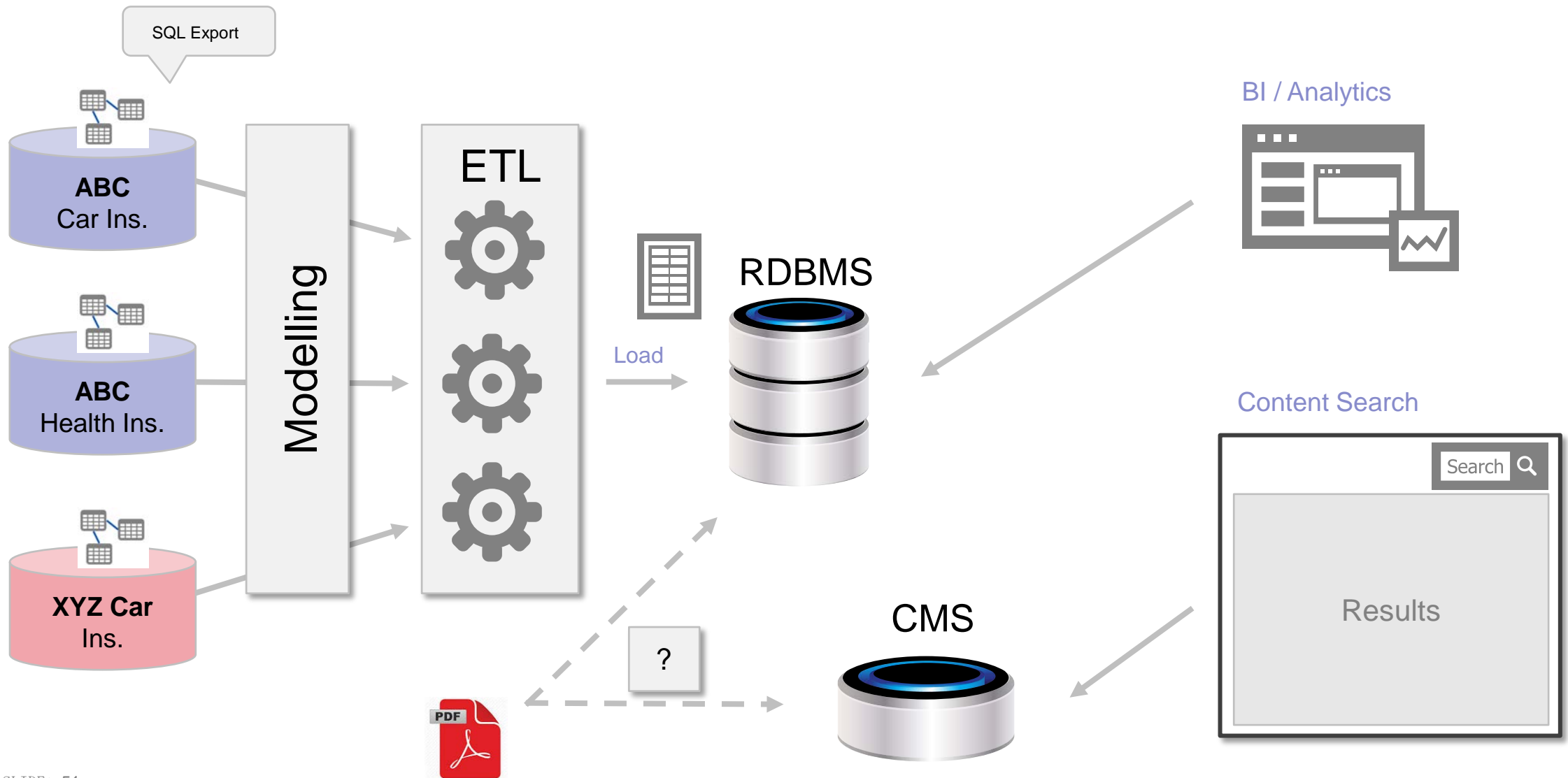
ABC Auto Insurance

ABC Health Insurance

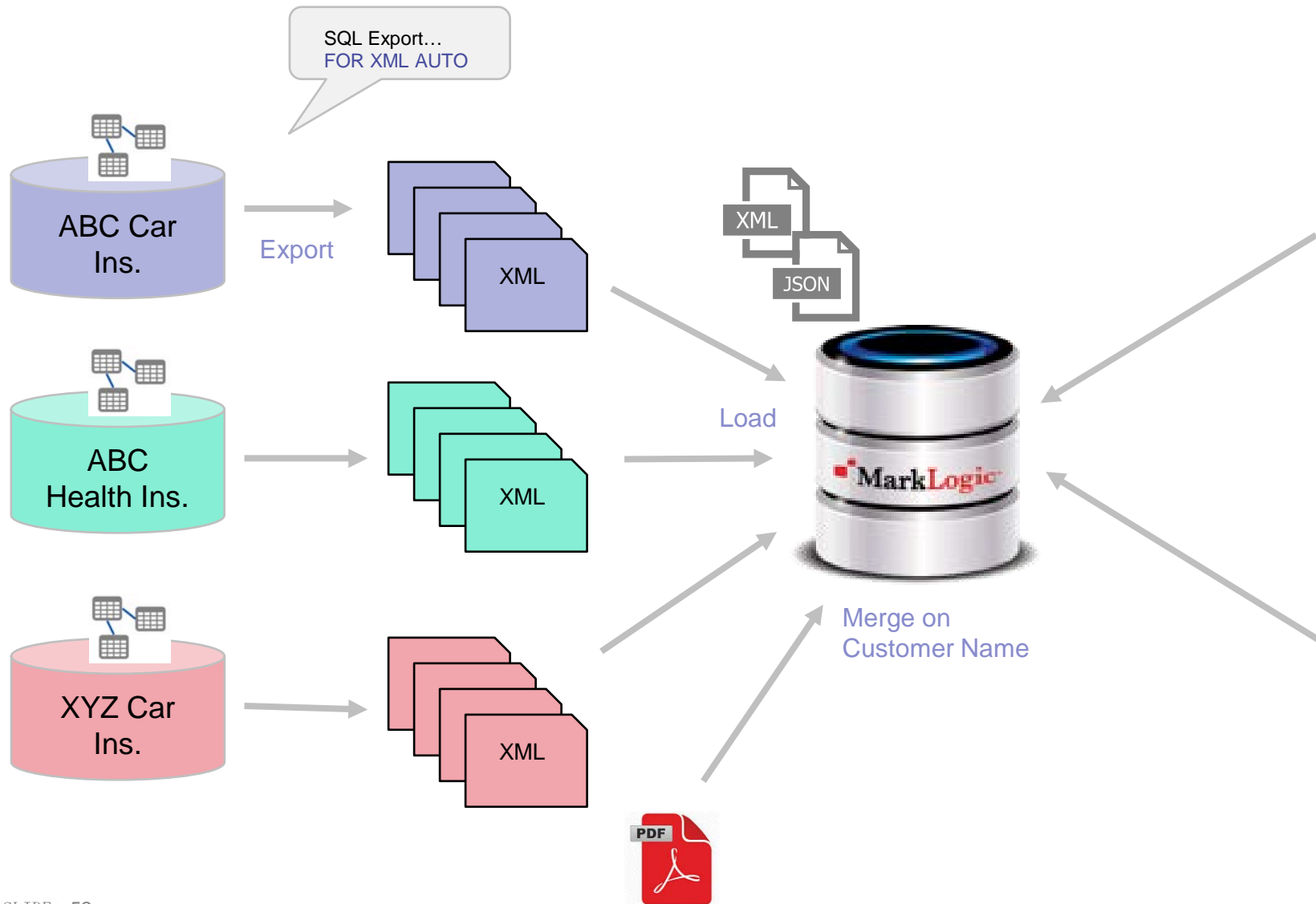
XYZ Auto Insurance



MarkLogic® リレーショナルモデルによるデータ統合



MarkLogic® ドキュメントモデルによるデータ統合



Search application with analytics



ODBC BI Reporting



```
<?xml version="1.0" encoding="UTF-8"?>
<policy-document>
  <policy>
    <Policy>AJS77HUI8NZ</Policy>
    <Start_Date>21/10/2015</Start_Date>
    <Billing_Term>Monthly</Billing_Term>
    <Yearly_Premium> $613 </Yearly_Premium>
    <Cust_ID>1000047</Cust_ID>
    <policy_rating>Investigation</policy_rating>
    <title>Policy ID: AJS77HUI8NZ Schneider</title>
    <company>ABC</company>
    <insurancetype>vehicle</insurancetype>
  </policy>
  <customer>
    <Cust_ID>1000047</Cust_ID>
    <Name>Griffith Schneider</Name>
```

```
<Cust_ID>1000047</Cust_ID>  
<policy_rating>Investigation</policy_rating>  
<title>Policy ID: AJS77HUI8NZ Schneider</title>
```

AND/OR

```
<Cust_ID>1000047</Cust_ID>  
<policy_status>Investigation</policy_status>  
<title>Policy ID: AJS77HUI8NZ Schneider</title>
```

エレメントの
名前を変える

AND/OR

```
<Cust_ID>1000047</Cust_ID>  
<policy_rating>Investigation</policy_rating>  
<policy_status>Investigation</policy_status>  
<title>Policy ID: AJS77HUI8NZ Schneider</title>
```

エレメントを
追加する

Configure MarkLogic

```
<policy_status> = <policy_rating>
```

MarkLogicの
機能を使う

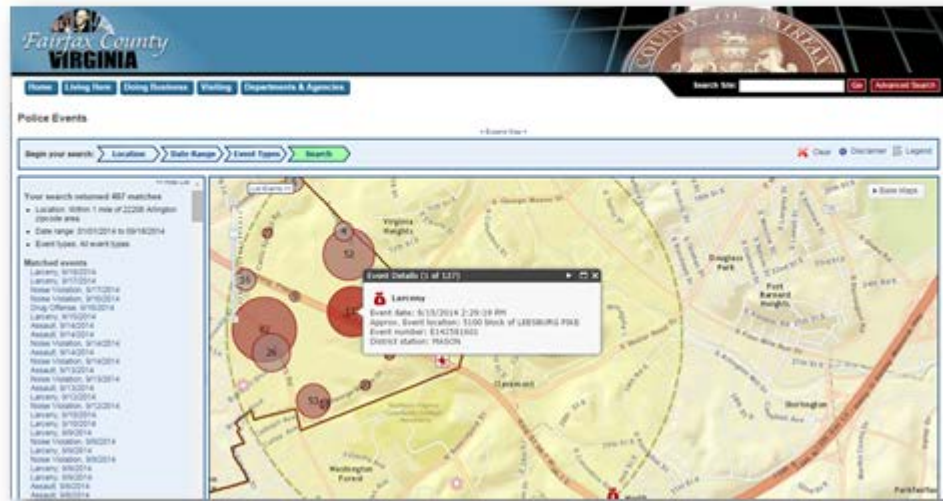
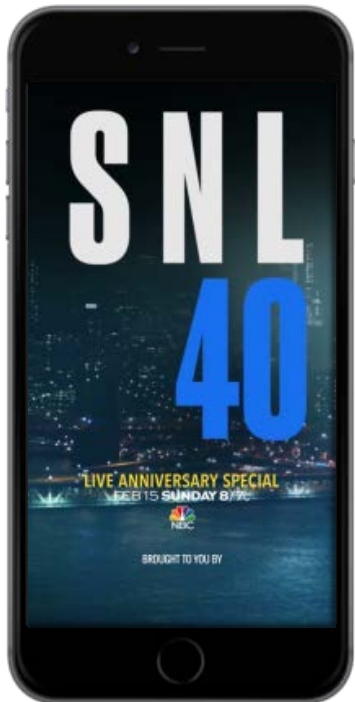
今日の問題

まとめ

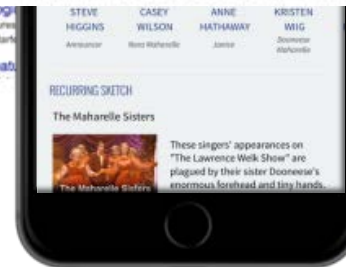
素晴らしいアプリケーションの実現



Fairfax County – Mapping Crime Data



SLIDE 22



ic Corporation

Website Directions

by Rd #200, San Carlos, CA 94070
100

Write a review Add a photo

arch for



© COPYRIGHT 2016 MARKLOGIC CORPORATION. ALL RIGHTS RESERVED.

Q&A