



知識検索Webサービス (日立製作所 + 日立システム)

2005年 6月10日

株式会社 日立システムアンドサービス
村垣 委久夫



知識処理技術

- 知識処理フレームワーク
 - データモデル (論理構造)
 - 知識処理手順 (作成時、応用時)
 - テキストマイニング技術を背景に持つ
 - 今回の知識検索サービスの基盤
- 百科知識
 - 百科事典コンテンツをもとに、知識処理フレームワークに基づき構築された知識ベース
 - 平凡社「世界大百科事典」、新聞、図鑑などとの連携を強化したコンテンツサービスに応用 (ネットで百科)



百科知識の内容



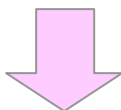
- ✦ 百科事典コンテンツの記述を基にした情報集合
- ✦ 概念を定義づける情報 (語句、数値) からなる
- ✦ 人物 / 地名 / 生物 / 建物に関する知識
- ✦ 経緯度、年月などの数値情報
- ✦ 複数の概念間の関係性と、その関係性を表す情報もあわせ持つ
- ✦ ある事柄と強く結びついた他の事柄
- ✦ 織田信長 本能寺の変、明智光秀、安土城、etc.



コンテンツサービスと知識処理

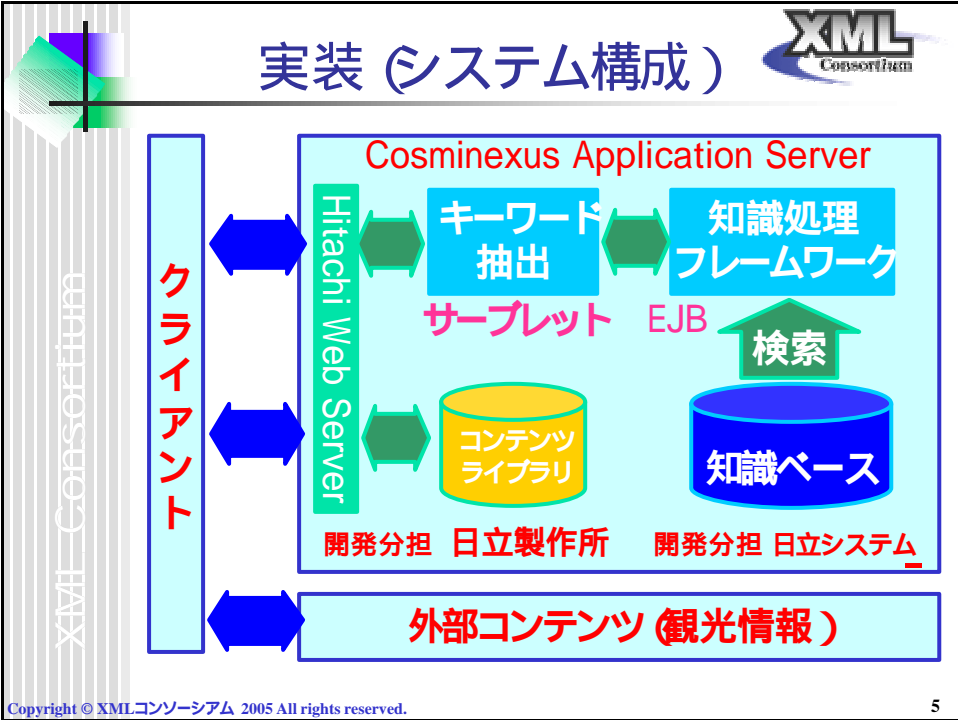


弊社 (日立システム) で行っているコンテンツビジネスの要素技術として活用

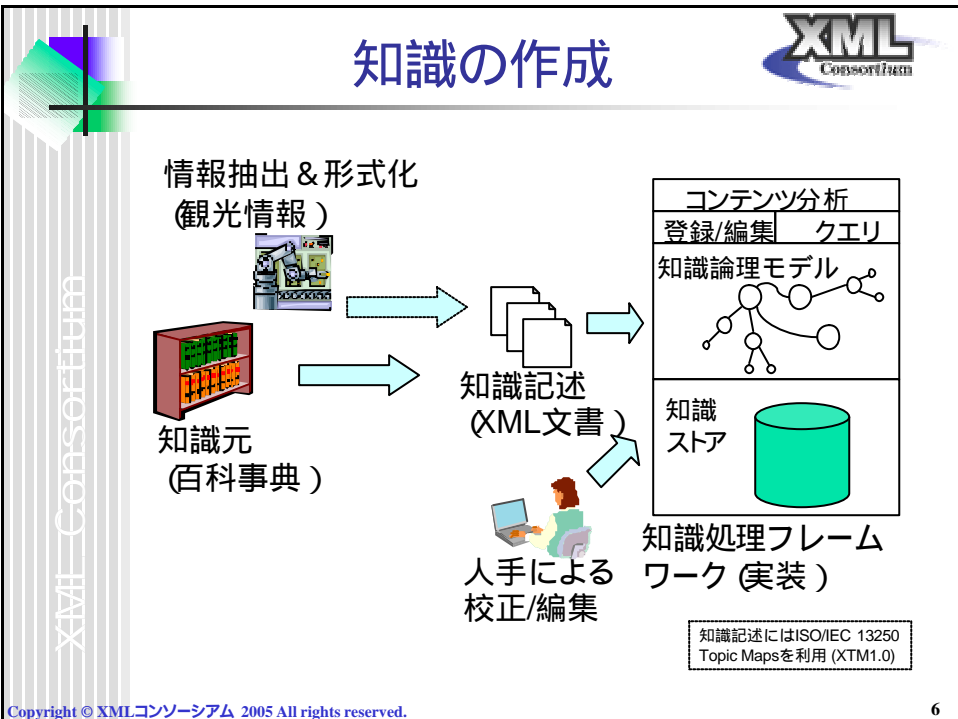


コンテンツサービスにおける付加価値創造
大量なコンテンツに対する付与作業を自動化
関連するコンテンツ間をたどるナビゲーションの提供
(関連情報へのリンク、検索)

実装 (システム構成)



知識の作成





知識検索サービスの機能



- 入力された文章に含まれるキーワードを抽出し、その検索キーワードに関連するコンテンツへのリンクを返す
 - **【入力】**コンテンツ説明文 :文字列
 - 処理の前段で検索キーワードを抽出
 - **【出力】**コンテンツへのリンク集合 URI リスト
 - 観光情報コンテンツ + (百科事典コンテンツ)



実証実験の成果



- 既存で作られた観光情報コンテンツの情報と百科辞典の知識ベースを連携した知識検索機能を効率よく構築した。
- 知識検索機能をWebサービスとしてサービス提供し連携することが有効な手段であることを実証できた。

今回の課題 (1)



- 文書検索機能との違いを表し難かった。
 - 知識検索のレベルを深くすると検索に時間が掛かりレスポンスが悪くなる。
 - 知識検索のレベルを浅くすると検索結果がキーワード項目に近くなり知識検索の効果が反映され難い。
- レスポンスと検索結果のバランスを調整し最適化を図るロジックを考える必要があった。

それは・・・



- たとえば

名古屋城

・徳川家康が建てた・・・

江戸城

・徳川家康が建てた・・・

知識検索の
結果として
特徴的

日光東照宮

・徳川家康が建てた・・・



今回の課題 (2)



- 動的なコンテンツ情報の収集・蓄積
 - 観光情報等日々情報が増えたり更新するコンテンツは知識を自動で収集する必要がある。
 - 自動収集した知識を定期的に見直し、編集する作業が必要。(雑音の除去)
 - 今回は期間限定であり 収集しないことにした
 - 今回は対象外としたがBlog情報など更新頻度の高い物では必須機能。



反省事項



- キーワード抽出機能追加への考慮不足
 - コンテンツの説明文から検索キーワードを抽出するようにしたが、説明文が長いと抽出されるキーワードが多くなり、検索の時間が長くなる。
 - 検索するキーワード数を制限したが、検索結果として得られるコンテンツ数が少なくなり知識検索の効果が半減した。
- 検索キーワード数を制限するのではなく、重要性で絞る方法を考えるべきであった。
- レスポンス性能とのせめぎ合いをどう解決するか。