



## iMAGE ～情報の再構成が生み出す意外な発見～

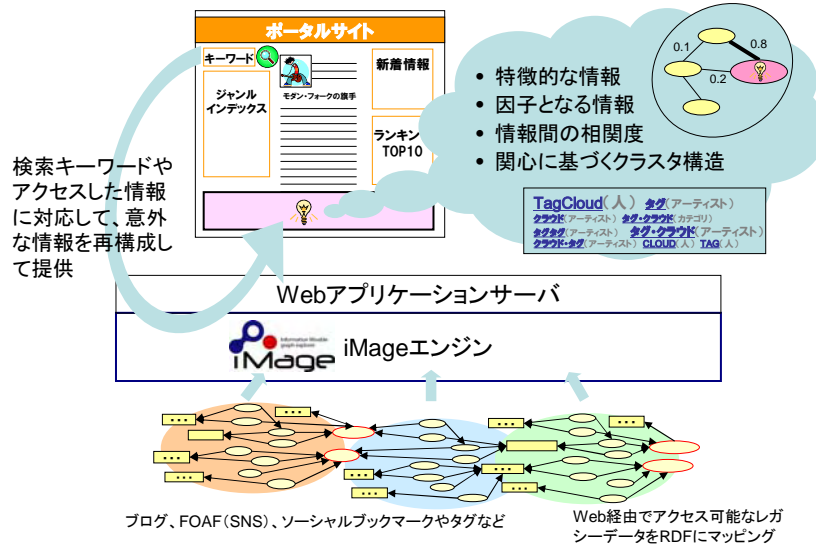
日本電信電話株式会社  
NTT情報流通プラットフォーム研究所  
佐藤 宏之 飯塚 京士 イコプラムディオノ  
大友 健治 村山 隆彦



## iMAGEの狙い

- 多様なノードとアークから構成されるグラフ構造データからの知識発見
  - 対象となるメタデータはWeb2.0により増加
- 異なる観点からデータに潜む知識を抽出
  - 特徴的な情報、因子、相関、クラスタ

## iMAGEの利用イメージ



(c)2007 NTT Information Sharing Platform Laboratories. All rights reserved.

3

## ブログと楽曲メタデータから知識発見

- 本発表では「ブログから抽出したメタ情報」や「楽曲メタデータ」を対象として解析
  - Doblog<sup>※1</sup>およびDoblogMusic<sup>※2</sup>のデータを利用
- アーティストや楽曲、ブログに関する特徴的な情報やその重要性、相関度などを抽出



※1 NTTデータが提供する無料のBlog(ブログ)レンタルサービスです。  
<http://www.doblog.com/>

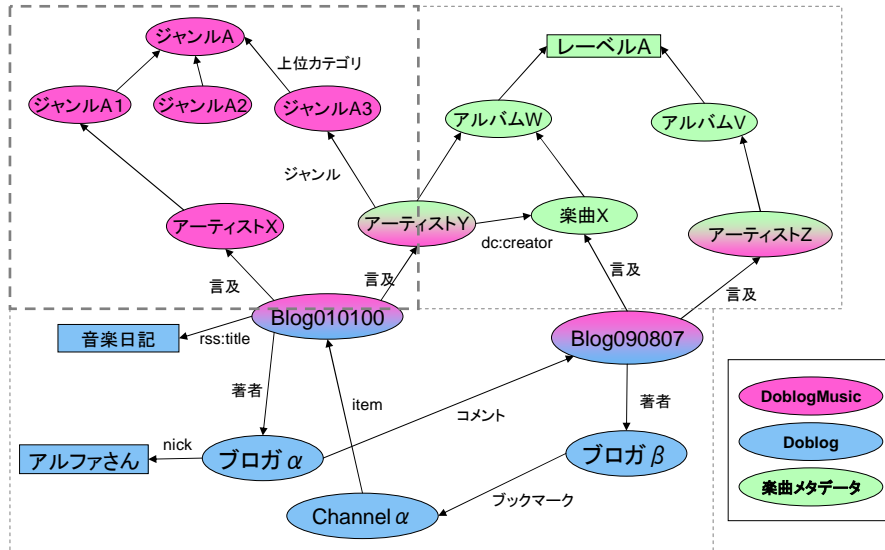


※2 NTTデータおよびNTT(ネットワークサービスシステム研究所)によるブログ記事から自動的にユーザの興味を判断しリコメンデーションや新たなコミュニティの創造支援を行う実験サービスです。  
<http://music.doblog.com/>

(c)2007 NTT Information Sharing Platform Laboratories. All rights reserved.

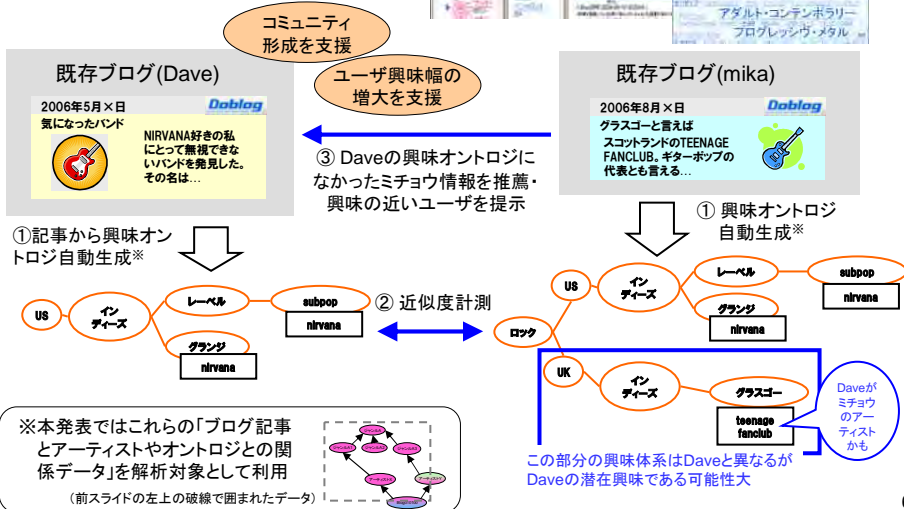
4

### 対象データのイメージ (3種類の異なるDB内容をグラフ構造で結合)



(c)2007 NTT Information Sharing Platform Laboratories. All rights reserved.

### (参考) DoblogMusicとは



※本発表ではこれらの「ブログ記事とアーティストやオントロジとの関係データ」を解析対象として利用  
(前スライドの左上の破線で囲まれたデータ)

## iMAGEによる特徴的な情報の検索



エリック・クラプトンに  
詳しくな人は誰？

### ■ ユーザの入力

キーワード:

ターゲット:

### ■ 結果

**記事の観点**  
(エリック・クラプトンに関する記事を2つ以上書いた人)

[Iiduka](#) さん  
[HSat](#)さん  
[OtomoK](#) さん ...

**コミュニティの観点**  
(記事を書くだけでなく他の人のクラプトンの記事にコメントしている人)

[IMura](#) さん  
[OtomoK](#) さん...

**周辺知識の観点**  
(記事を書くだけでなく同じジャンルの他のアーティストの記事も書いている人)

[IkoP](#) さん  
[HSat](#) さん...

クリックして  
ブログ  
記事を  
表示

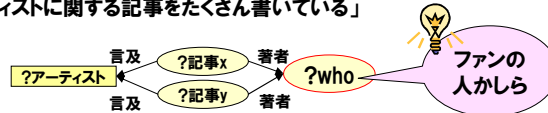
(c)2007 NTT Information Sharing Platform Laboratories. All rights reserved.

※ユーザ(ブログ)名は仮名に変更しています。 7

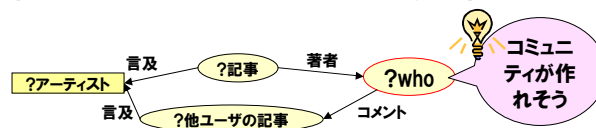
## なぜ、観点毎の結果が得られるのか？

- 2つの概念が深い関係にある際の因子構造(頻出パターン)を解析し、そのパターンに合致した対象のみを検索

「アーティストに関する記事をたくさん書いている」



「記事を書くだけでなく、同じアーティストに関する他の人の記事にコメントしている」



(c)2007 NTT Information Sharing Platform Laboratories. All rights reserved.

8

# こんな検索も



エリック・クラプトンに関する曲は？

## ■ ユーザの入力

キーワード:

ターゲット:

## ■ 結果

これは従来の検索でも出せる

(クラプトンが書いた / 演奏した曲)  
**HIGH**  
 いとしのレイラ  
 I want a little girl  
 After midnight...

(クラプトンの記事を書くブロガが言及する他のアーティストの曲)  
**ベニーとジェッツ (エルトン・ジョン)**  
**無法の世界 (ザ・フー)**  
**Do They Know It's Christmas (バンド・エイド)**  
**オールド・ブラウン・シュー (ジョージ・ハリスン)**

クラプトン度が高い曲？

# クエリパターンの再利用

記事の観点  
 (エリック・クラプトンに関する記事を2つ以上書いた人)

- liduka さん
- HSat さん**
- Otomok さん ...

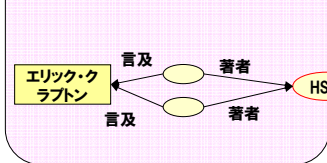


この人が詳しい他のアーティストは？

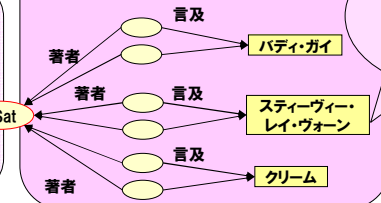
関連 クリック

## ■ 結果

### 最初の検索結果



### 関連



スティーヴィー・レイ・ヴォーンやクリームについても同じように詳しいのか

## 因子ノードの抽出



手っ取り早く、クラブトンに関連の深いものを知るには？

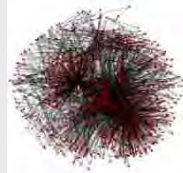
### ■ ユーザの入力

キーワード: **エリック・クラブトン**

因子

クリック

「エリック・クラブトン」に  
関係する頻出パターンに合  
致するサブグラフを抽出  
してネットワーク分析した



抽出したサブグラフ

### ■ 結果

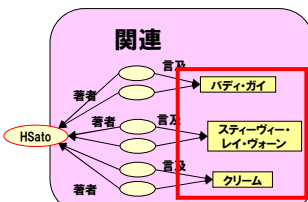
**Otomok(人) マドンナ(アーティスト)**  
 マリア・キャリー(アーティスト) アダルト・コンテンポラリー(カテゴリ)  
 ステイング(アーティスト) ブルース・スプリングスティーン(アーティスト)

**U2(アーティスト) マドンナ(アーティスト)**  
 ザ・フー(アーティスト) クイーン(アーティスト)  
 エルトン・ジョン(アーティスト) バンド・エイド(アーティスト)  
 ジョージ・ハリスン(アーティスト) DVD「ライブ・エイド」(ブログ記事)

曲に関する  
サブグラフか  
ら抽出

ブログに関する  
サブグラフから  
抽出

## 関係抽出

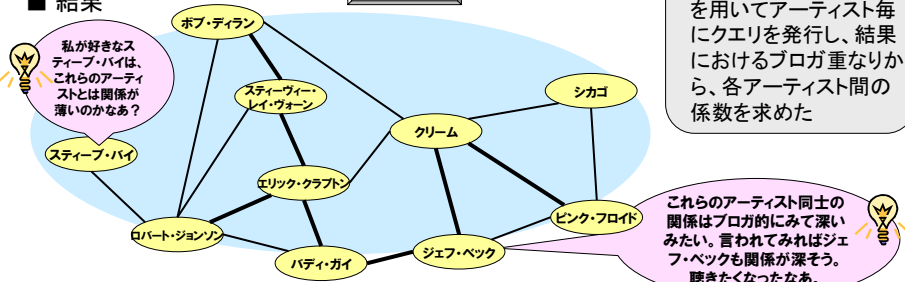


これらのアーティスト同士の  
関係や、自分の興味あ  
るアーティストとの関係は  
わからないの？

クリック

相関度

### ■ 結果



私が好きなス  
ティーブ・バイは、  
これらのアーティ  
ストとは関係が  
薄いかなあ？

記事の観点(パターン)  
を用いてアーティスト毎  
にクエリを発行し、結果  
におけるブログ重なりか  
ら、各アーティスト間の  
係数を求めた

これらのアーティスト同士の  
関係はプロガ的にみて深い  
みたい。言われてみればジェ  
フ・ベックも関係が深そう。  
聴きたくなったなあ。

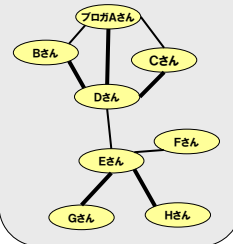
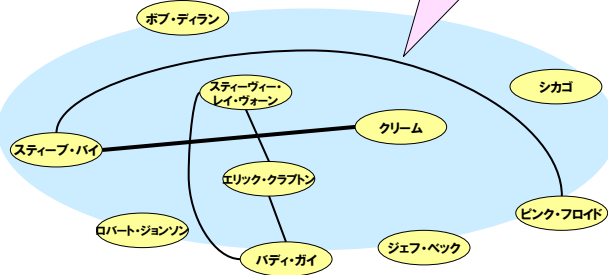
# 観点が異なると関係も変わる

アーティストが所属するジャンルの重なりによる各アーティスト間の係数を求めた

※1つのアーティストが複数のジャンルにまたがって所属することも多い

厳密なジャンルのつながりはあまりないね。でも見えていなかった関係が見えた。記事の観点と比較すると面白い。

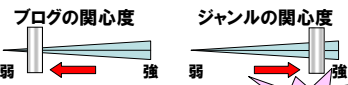
同様の方法で観点毎に異なるプロガ同士の関係も抽出できる



# クラスタ抽出

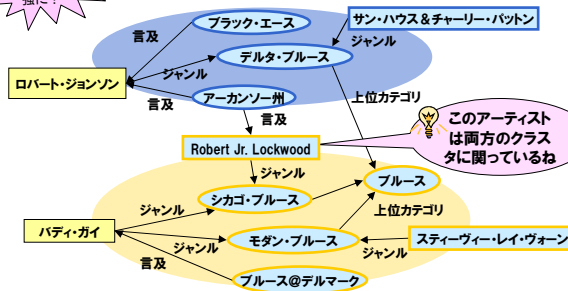
■ ユーザの入力

キーワード:



これらのアーティストとジャンルの近いアーティスト達は？

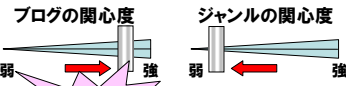
対象データのアーキの種類の違いによる重みも考慮したページランクのようなアルゴリズムを用いて、ランクの高いノードのクラスタを抽出



## 関心が異なるとクラスタも変わる

### ■ ユーザの入力

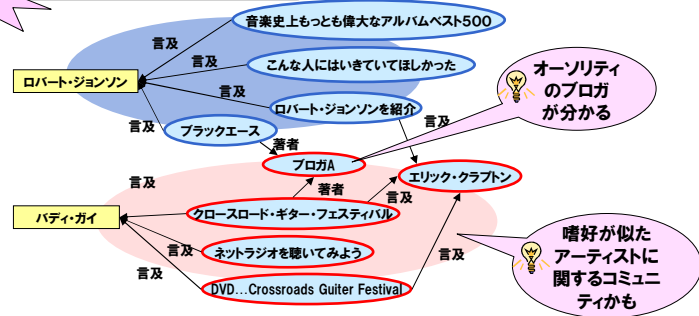
キーワード: **ロバート・ジョンソン バディ・ガイ**



今度はこっちを強に!



ブログの関心度を強くしたら、どんなクラスタができるのだろう?



オーソリティのブログが分かる

嗜好が似たアーティストに関するコミュニティかも

## おわりに

- iMAGEは異なるデータソースから構成された多様なノードとアークを含むグラフ構造データ (RDF) に適用可能
- 解析結果の知識表現パターン (観点) を活用し、異なる観点からさまざまなタイプの情報を抽出
- 企業内のデータを対象とした実験も実施中