

# Folksonomyによる 階層構造画像データベースの 構築

電気通信大学 情報工学専攻  
柳井研究室

秋間 雄太, 川久保 秀敏, 柳井 啓司  
2010年7月28日

# Folksonomyを用いた 画像特徴とタグ共起に基づく画像オントロジー の自動構築

(Folksonomyによる階層構造画像データベースの構築)

電気通信大学 情報工学専攻  
柳井研究室

秋間 雄太, 川久保 秀敏, 柳井 啓司  
2010年7月28日

# 背景

- Folksonomyとは？
  - ユーザによる分類

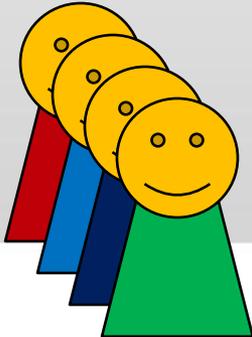
コンテンツ



利用



ユーザ



分類

一般ユーザが自由に  
タグ付けを行うことで  
ユーザ自身が扱いやすい  
データセットが構築される

# 背景

- Folksonomyの欠点

語彙が統一されていない！

メン

fall

語彙間の関係を  
構築することで解決



# 背景

Folksonomyに基づいた画像データセット

- オンラインアルバムサービスFlickr
- タグなどによって意味的な価値を画像に付与



しかし、

- 基本的に付与されているタグは単なる単語の羅列
- 概念間の関係を加味したものではない



Folksonomyの意味情報と、画像の視覚情報を用いて、  
オントロジーが作成できれば、より詳細な画像検索などが可能になる

# オントロジーとは



検索エンジン  
機による  
関連付け  
メニュー

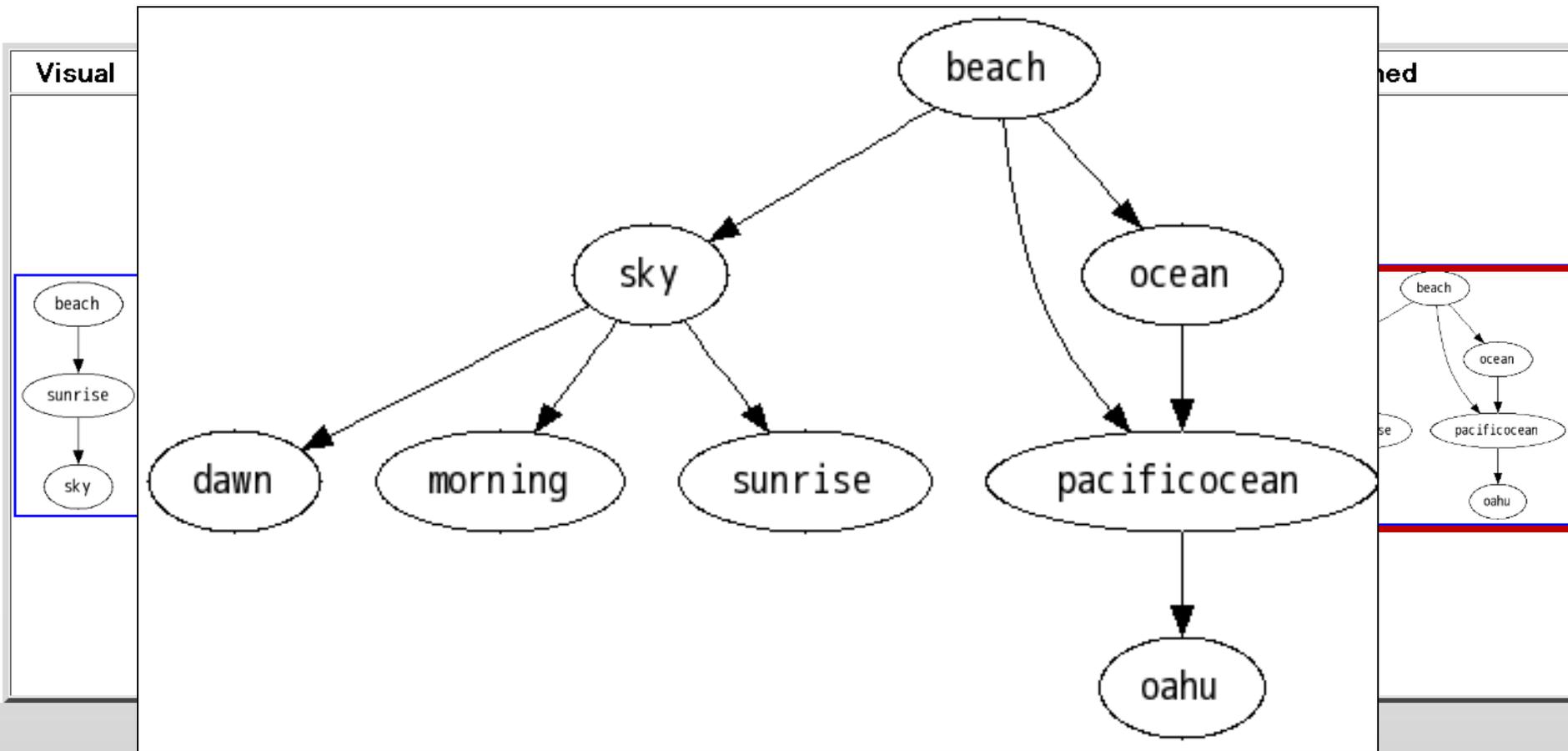
- 単なる画像検索から意味情報を考慮した検索へ
  - beachとしかタグが付いていない画像をseaでも検索可能に

# 目的・手法概要

## ■本研究の目的

- 階層構造を持った画像データベースの構築
  - 200万枚の画像からなる大規模画像データセットを用いる
  - 視覚による表現，タグによる表現，視覚とタグを統合した表現の3種類でそれぞれ構築
  - 特徴分布の広い概念ほど上位の概念であると仮定して上位下位関係を推定

# 得られる階層構造の例



# 階層構造自動構築の利点

- 画像検索の補助として利用可能
  - クエリ拡張に利用し、入力検索クエリに直接対応しない画像も検索対象に。
  - 語彙の関係性を利用した範囲の絞り込み
- 画像認識や画像アノテーションへの利用
  - 画像データセットは各概念について画像特徴量の分布情報を保持

# 関連研究

## ■ WordNet

- 概念間の意味的関係を利用
- 上位下位関係
- 画像検索
- 固有名詞

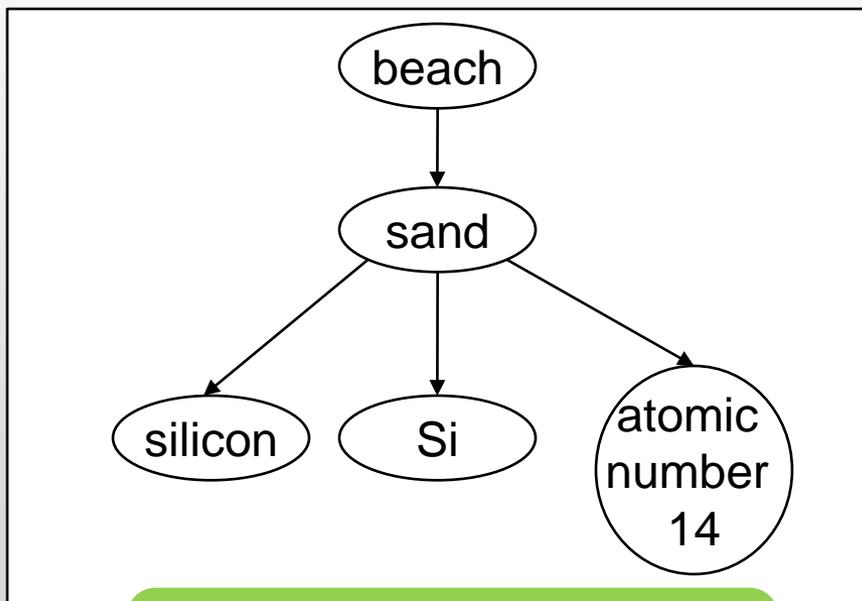
視覚情報による  
自動的に作成された  
階層構造を  
オントロジーとするような  
研究は少ない

## ■ 視覚

- Tangら  
IJCAI
- Plangprasopchokら  
Flickrの既存の関係を利用  
[Plangprasopchok et al ACM WWW Conf2009]

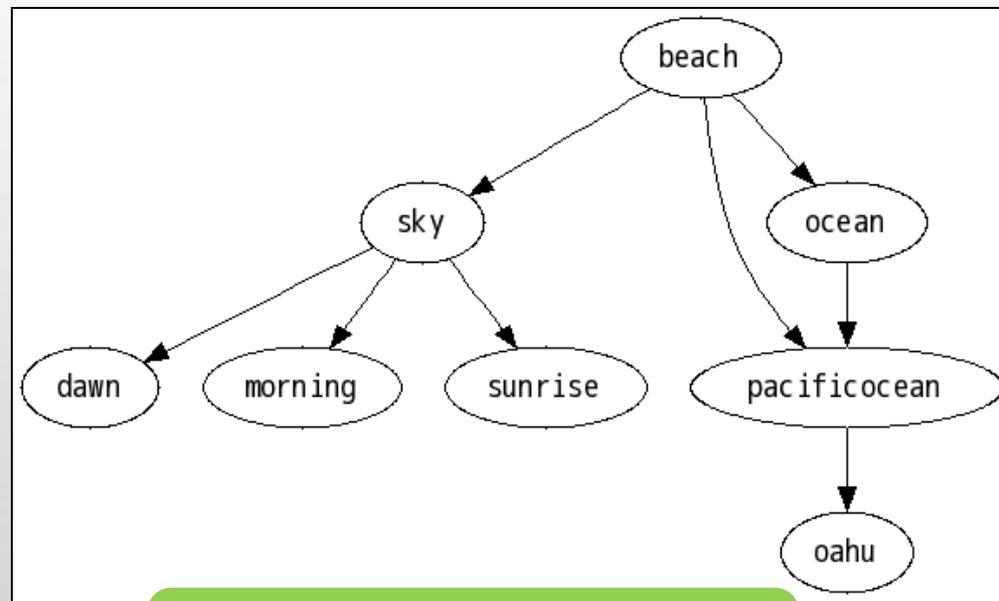
# WordNetとの比較例

## WordNetの階層の例



専門的な知識も

## 本実験で作成される階層の例



直観的な関係

# 階層構造構築の手順

## 大量画像取得と特徴抽出

- Flickrからおよそ200万枚のタグ付き画像を取得
- 画像特徴とタグ共起を用いた3種類の方法で画像を表現

## ノイズ画像除去

- pLSAを用いてノイズ画像を除去

## 概念ベクトルの作成

- 各概念を表現するトピックベクトルを生成

## 概念間距離の測定

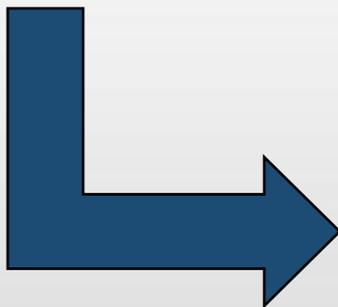
- 概念ベクトルから各概念間の距離を測定
- 距離によって関係する概念を取得

## 概念間階層構造の構築

- 概念ベクトルからエントロピーを算出
- エントロピーの大小で上下関係を決定

# Flickr APIでの画像取得について

ランダムに  
200万枚分の  
画像情報を取得



タグ

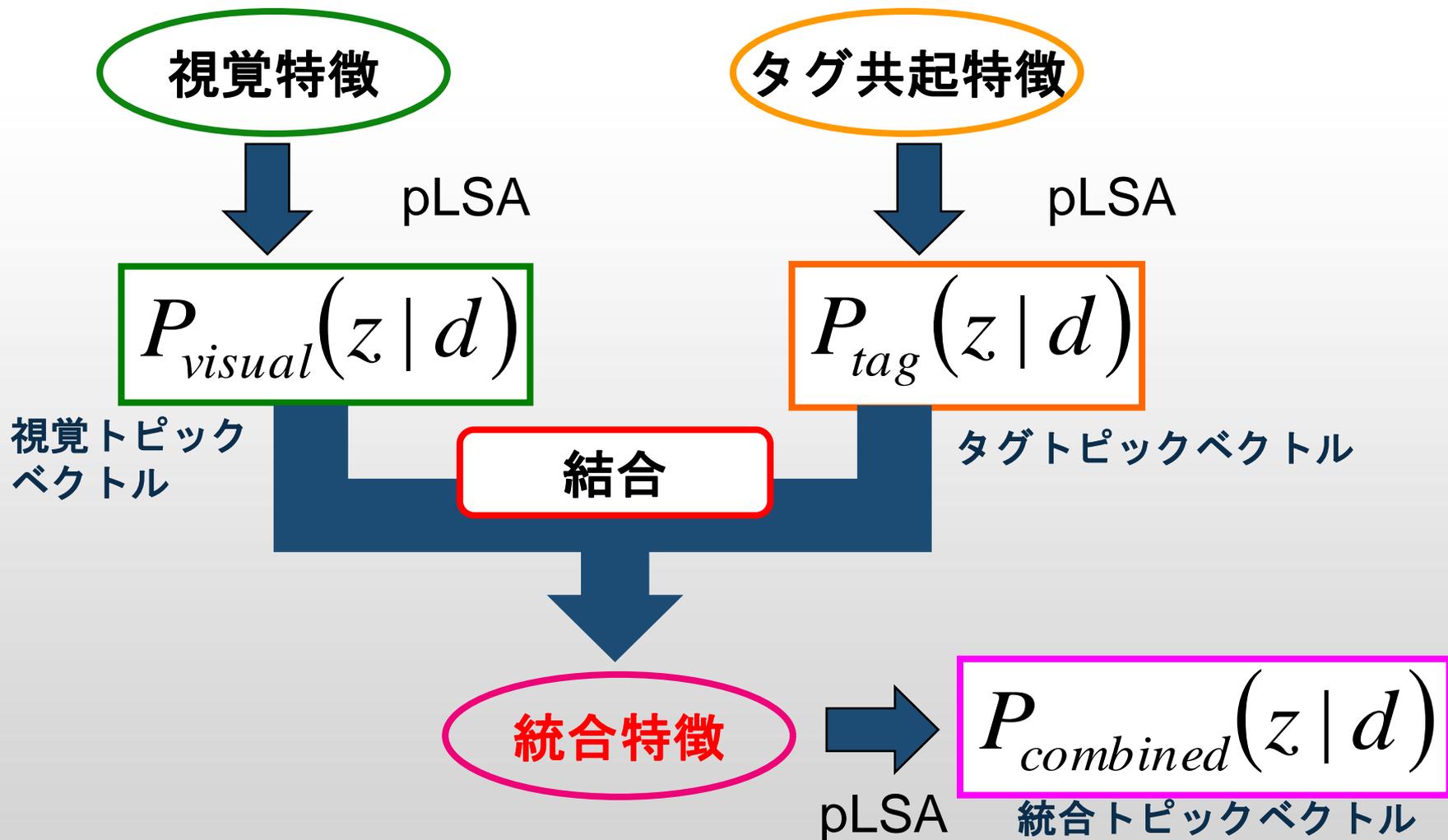
- bird
- goose
- swanny
- Bali Bird  
Park

取得できる画像例

# 画像の表現方法

- 視覚特徴による表現
  - Bag-of-Keypoints表現
  - ベクトル次元 : 1000
- タグ特徴による表現
  - Bag-of-Words表現(Bag-of-Tags表現)
  - ベクトル次元 : 4345

# トピックベクトルの算出

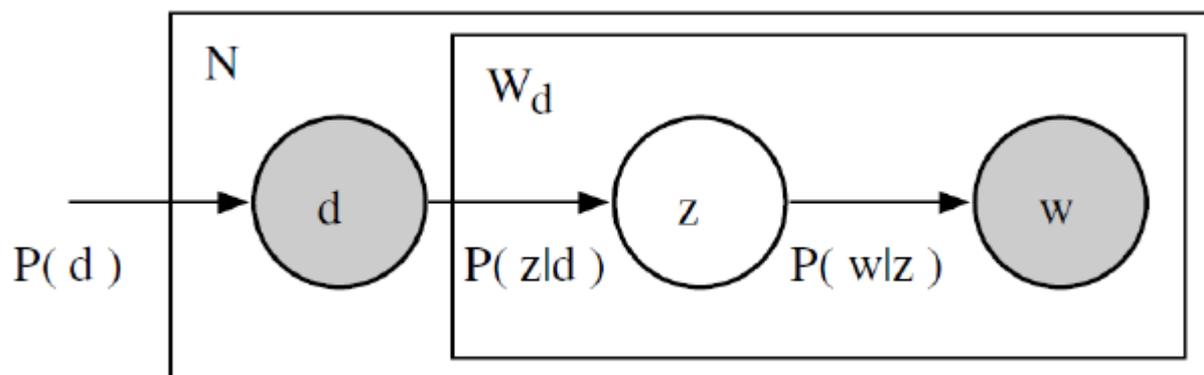


(各トピックベクトルは、実験では100次元とした)

# probabilistic Latent Semantic Analysis

- 確率的なクラスタリング手法
  - 画像は複数の潜在トピックへ確率的に帰属する
  - 確率的分類が最適になるような、  
画像  $d$ , 単語  $w$ , 潜在トピック  $z$ 間の関係を算出

$$P(d_i, w_j) = \sum_{k=1}^K P(d_i|z_k)P(w_j|z_k)P(z_k)$$



# ノイズ画像除去

$$P_{Concept}^{z_k} = \sum_{d \in Concept} P(d|z_k) / |Concept|$$

Concept : moon

PLMに与える  
トピック作成

重みが大きくなる

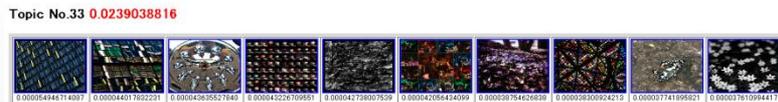
$$P(Concept|z_k) = \frac{P_{Concept}^{z_k}}{P_{Concept}^{z_k} + P_{NotConcept}^{z_k}}$$

トピックの概念らしさの  
度合い

Not moon dataset

画像と概念の関係性を計算

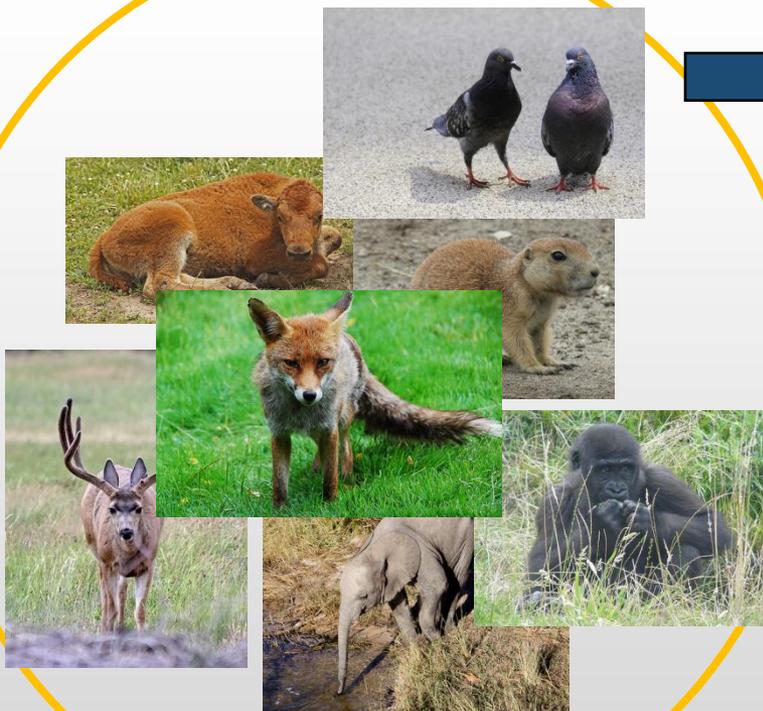
$$P_{NotConcept}^{z_k} = \sum_{d \in NotConcept} P(d|z_k) / |NotConcept|$$



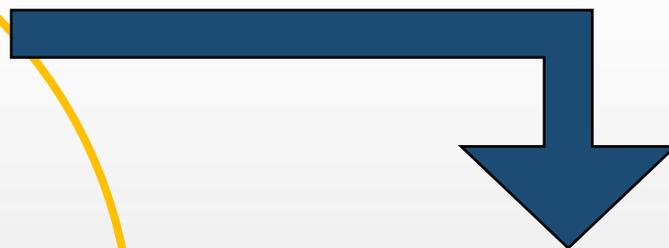
$$\sum_{k=1}^K P(Concept|z_k) P(z_k|d_i)$$

# 概念ベクトルの生成

$$P(z|Concept) = \left( \sum_{d \in Concept} p(z|d) \right) / (Concept \text{ の画像枚数})$$



Animal



Animal  
概念ベクトル

# 概念間関係

## ■ 概念間距離

- JSダイバージェンスを利用

- P, Q はそれぞれの概念の確率分布(ベクトル表現)

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

$$D_{JS}(P||Q) = \frac{D_{KL}(P||(P/2 + Q/2))}{2} + \frac{D_{KL}(Q||(Q/2 + P/2))}{2}$$

# 概念間関係

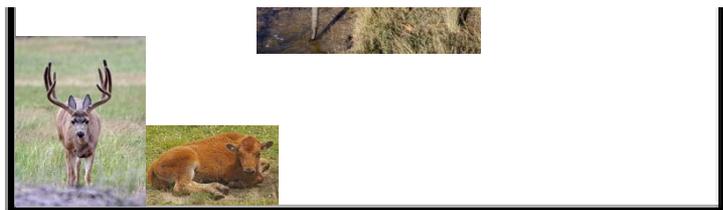
## ■ 概念間の上下関係

$$H(\text{Concept}) = - \sum_{z \in Z} P(z|\text{Concept}) \log(P(z|\text{Concept}))$$

上位概念 (ばらつき大)



エントロピー大

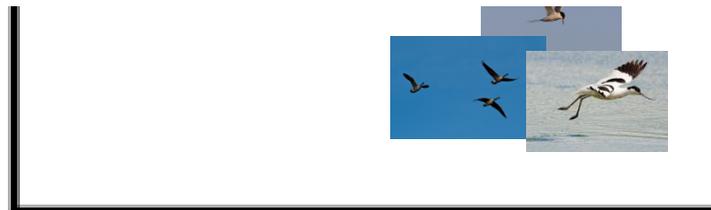


Animal

下位概念 (ばらつき小)



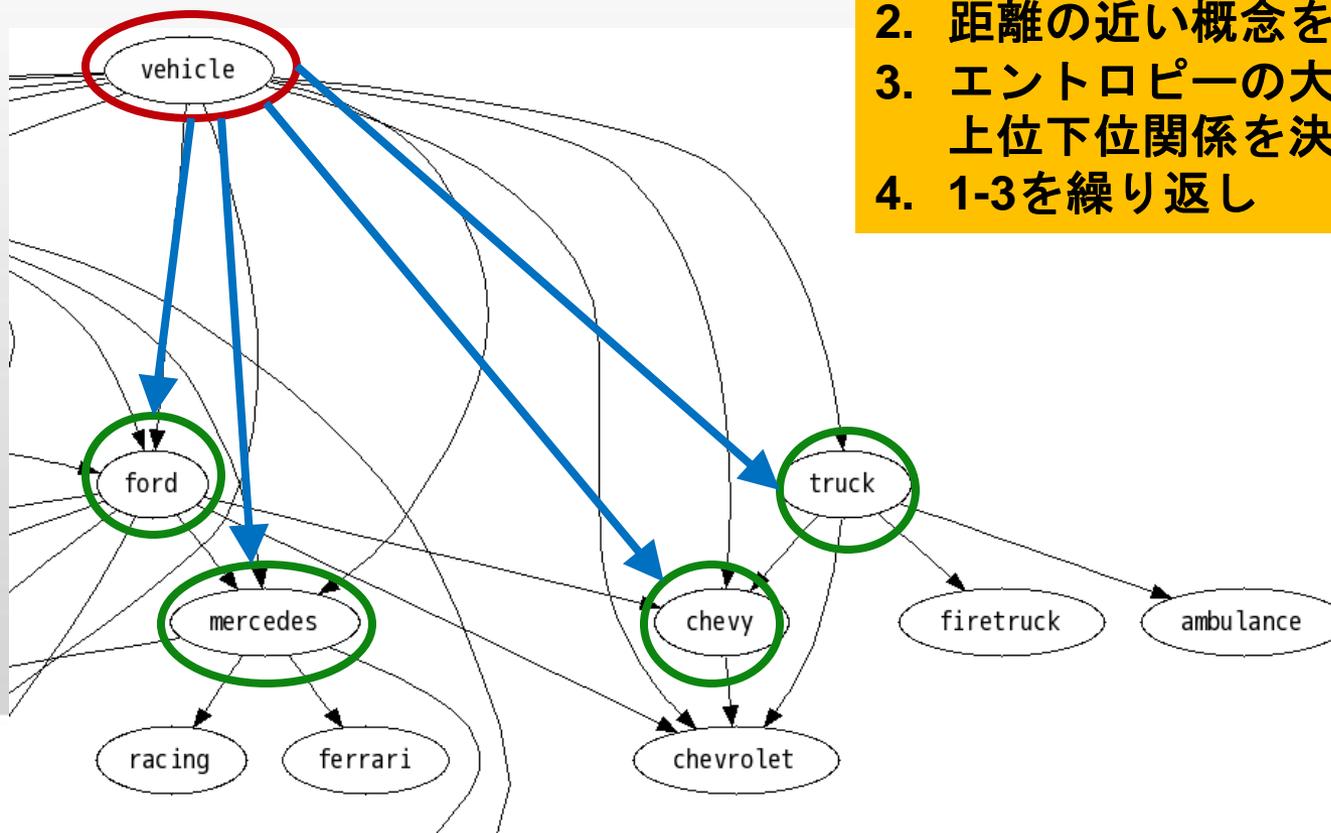
エントロピー小



Bird

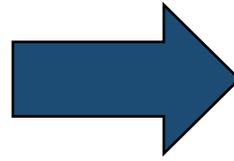
# 階層構造の構築方法

- 非巡回有向グラフ(Directed Acyclic Graph, DAG)による表現



1. 起点となる概念を決める
2. 距離の近い概念を算出する
3. エントロピーの大小によって上位下位関係を決定
4. 1-3を繰り返す

関係とは



road

本実験では  
これらのそれぞれの区別は  
行っていない  
(今後の課題)

red

adjective



# 実験について

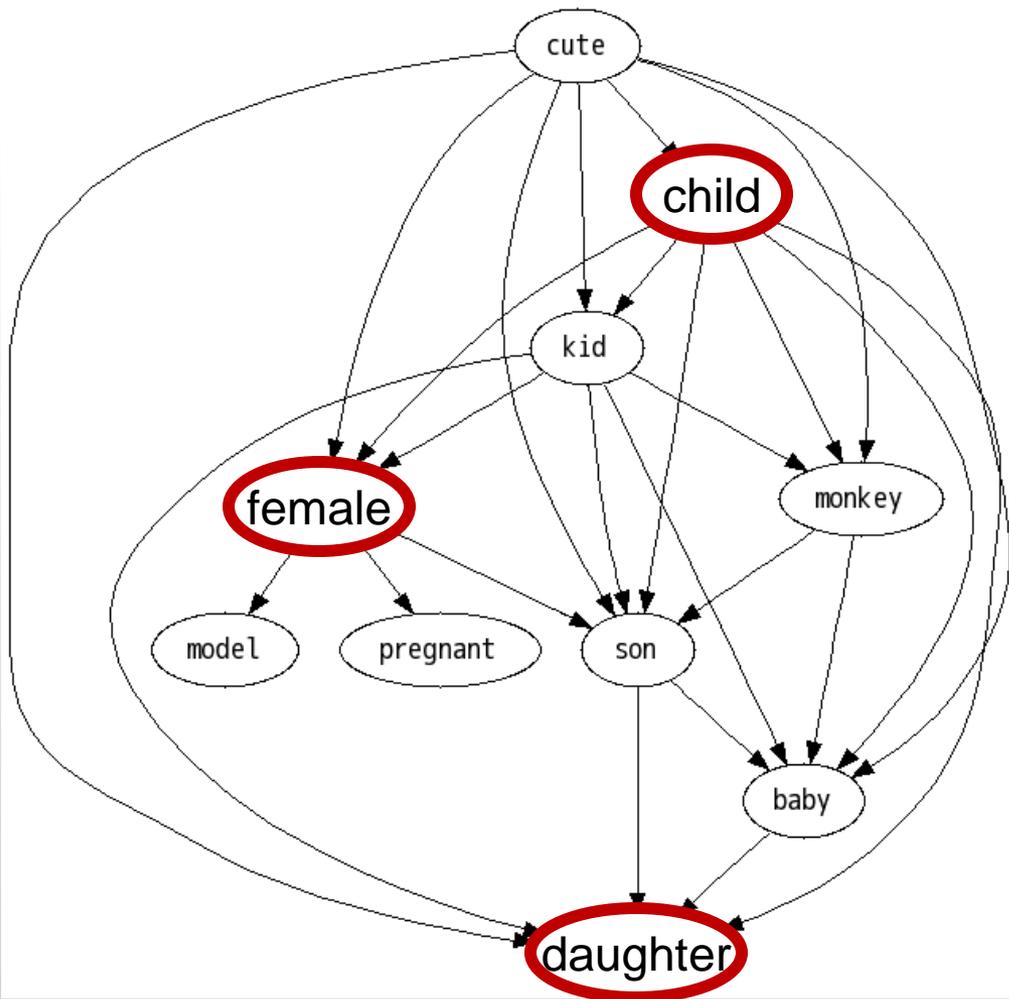
- Flickr画像による大規模画像データセット
  - 合計およそ**200万枚**のタグ付き画像
  - 階層構造を構築する概念の数は2657

- 各設定

- ノイズ除去によって、各概念の上位n枚は  
 $5 \times \sqrt{(\text{各概念の取得枚数})}$  を用いて計算する

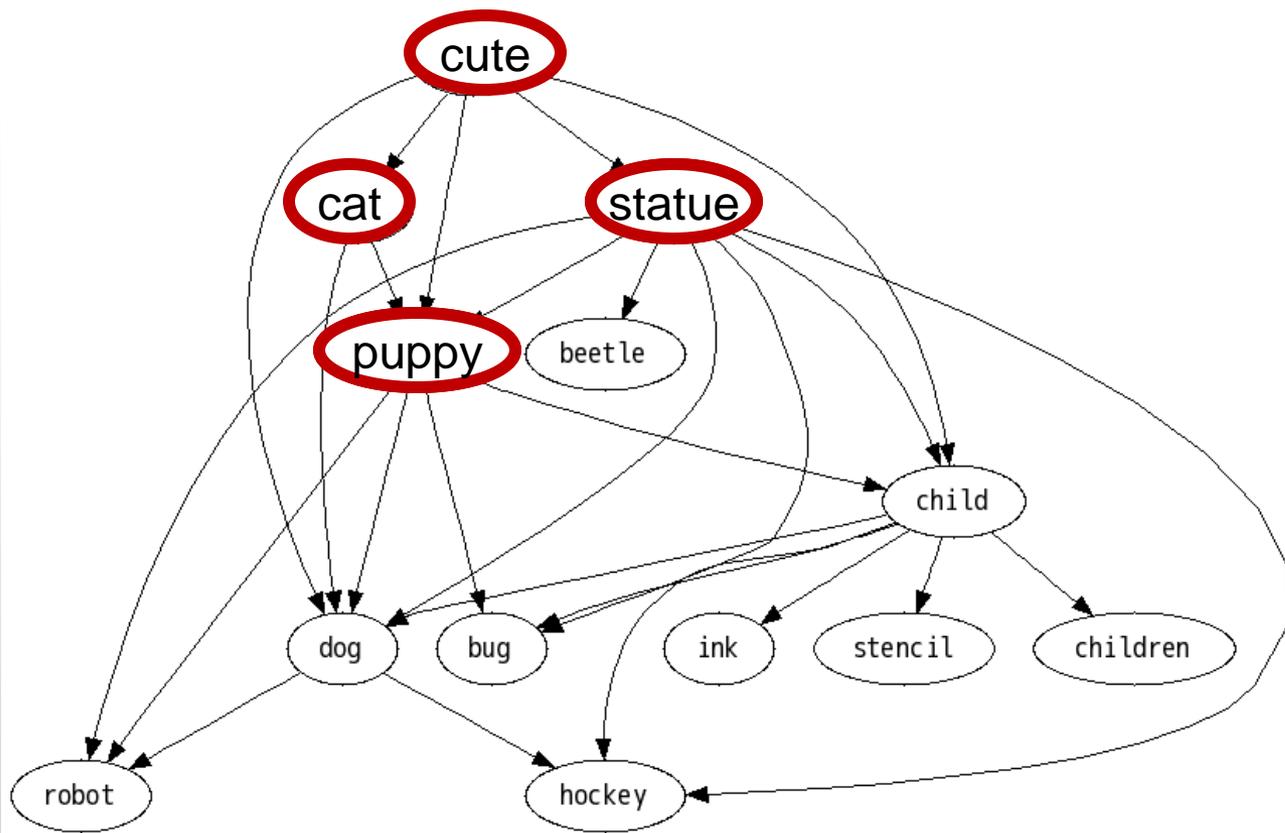
- これは画像枚数が少ない概念(200枚程度)でも  
およそ100枚の画像を利用できるように設定

# cuteの下位構造



タグ表現による  
階層構造

# cuteの下位構造

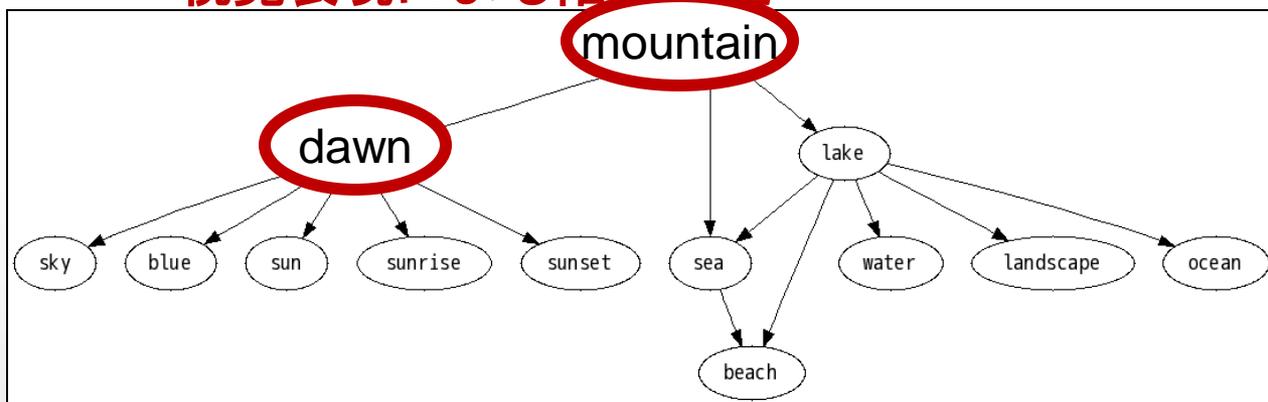


視覚表現による  
階層構造

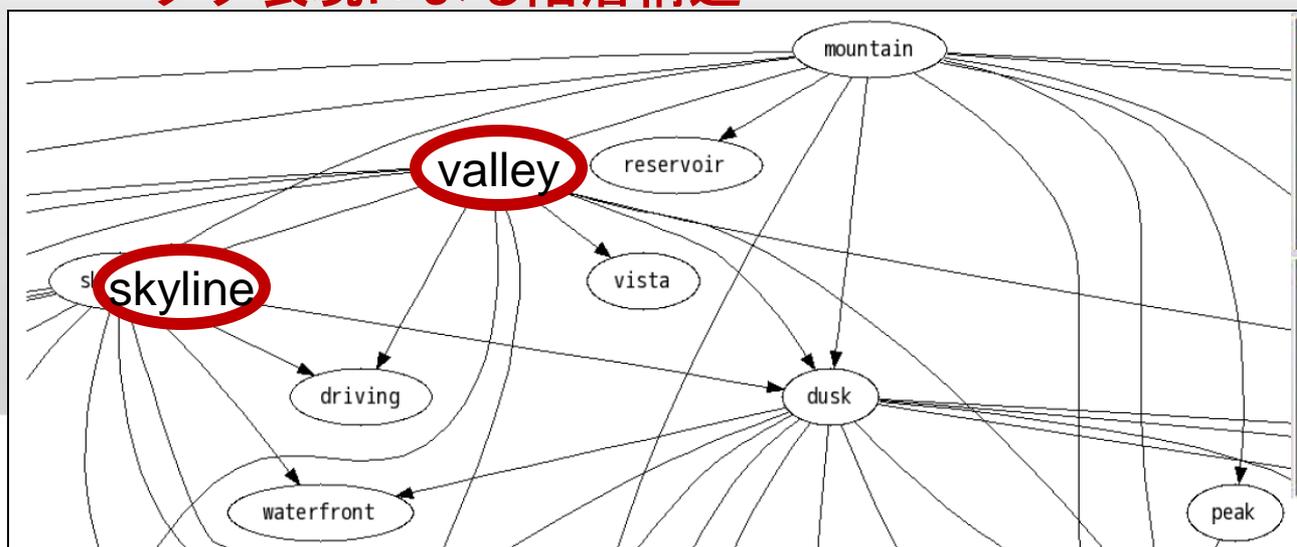


# mountainの下位構造

## 視覚表現による階層構造

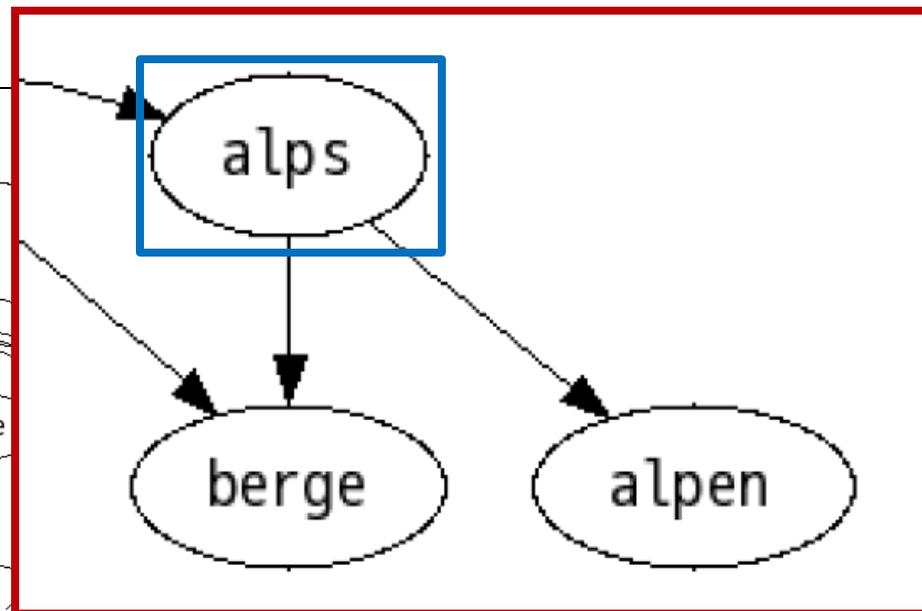
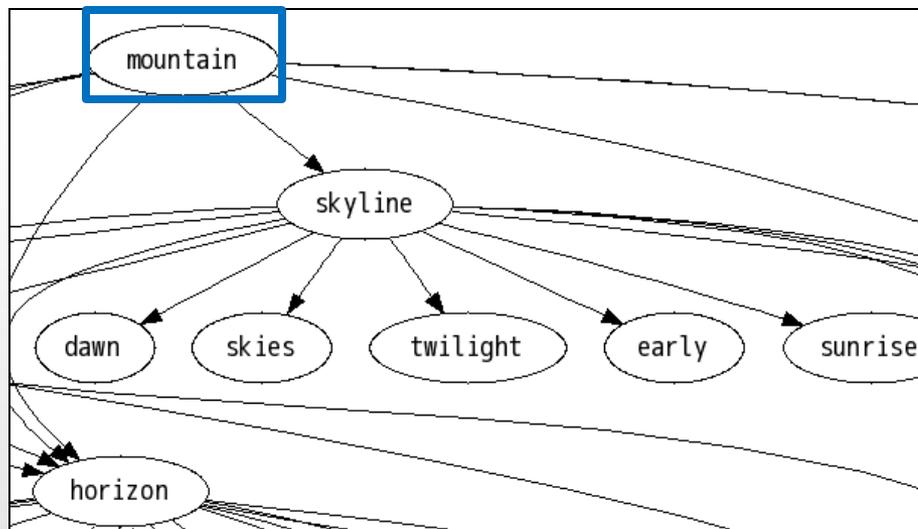


## タグ表現による階層構造



# mountainの下位構造

## 統合表現による階層構造



# 表現方法の比較

## ■表現方法ごとの違い

- **視覚表現**：視覚的な類似を持った関係性
  - データベース特有あるいは人間の知り得ない新しい関係性を抽出できる可能性がある
- **タグ表現**：意味的に関連する言葉としての関係に近い
  - 人間の認識に近い関係を抽出できる可能性がある
- **統合表現**：両方を足し合わせたような関係で画像検索などに利用できる可能性が高い

# 利用と評価のための予備実験

- Flickr上での検索で、直接の上位語・下位語をAND検索クエリとして追加
  - 上位10枚の精度を人手で評価
    - 元となるクエリに対応する画像かどうか
  - 100のクエリについて評価

	追加前	下位語追加	上位語追加
OK画像数平均	7.4枚	8.1枚	8.1枚

- 評価できないクエリはスキップ：外国の地名など
- 実験の再設計と、規模の拡大

# まとめ

- 階層構造を持った画像データベースの構築
- 構築した階層構造について
  - タグと視覚に基づき概念の関係を抽出
  - 表現方法の違いで多様な形を表す
  - 検索システムなどへの活用の可能性を発見

## 今後の課題

- 他の視覚特徴量も利用する（色など）
- 関係性の種類の認識を行えるようにする
- 既存のテクノロジーの活用
  - 例：Wordnet内にある単語では、そのデータを活用
- もっと大規模な評価を行う

ご清聴ありがとうございました

ポスター  
IS3-70