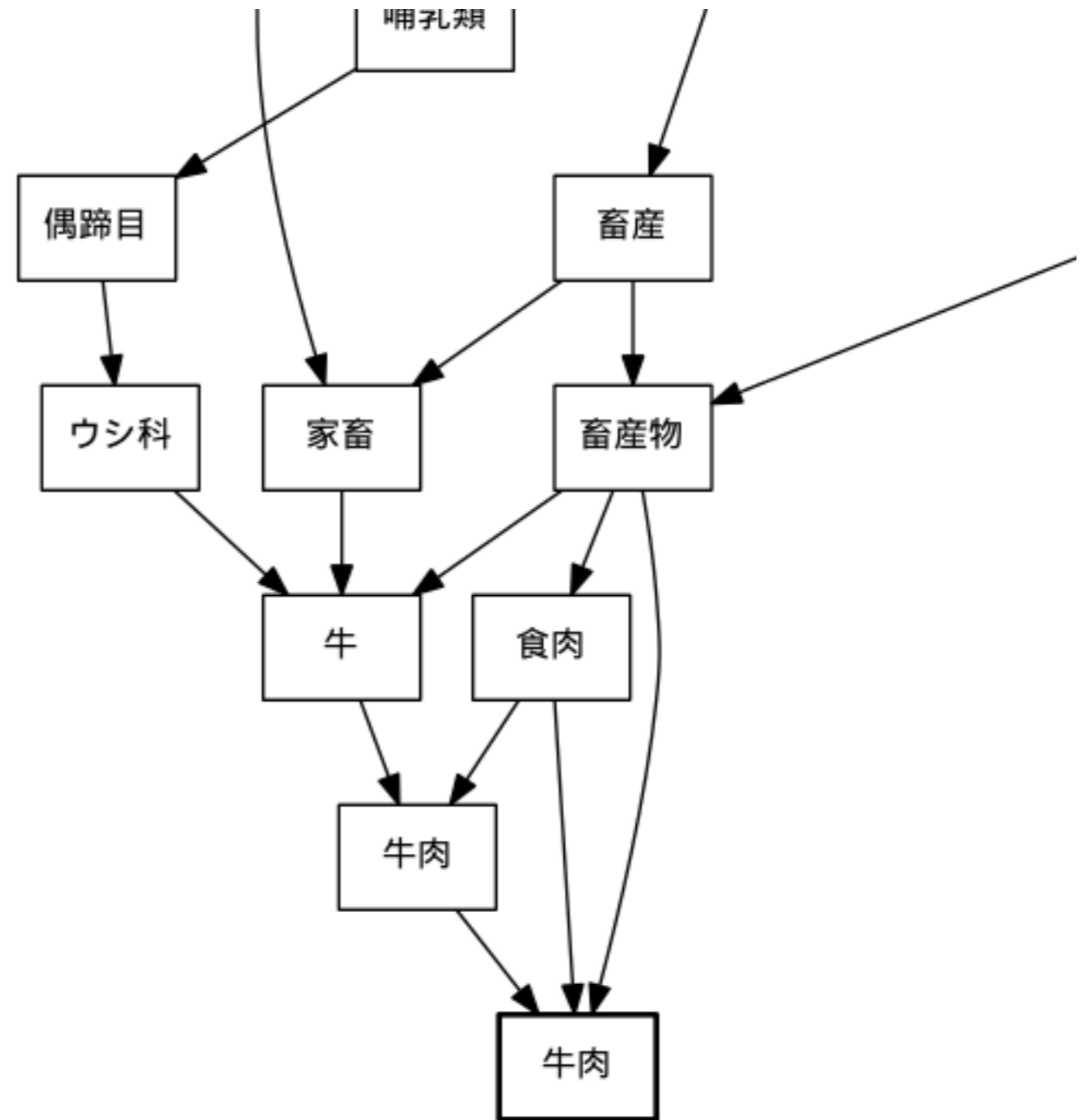


Wikipedia のための  
Latent Dirichlet Allocation  
小松 弘佳

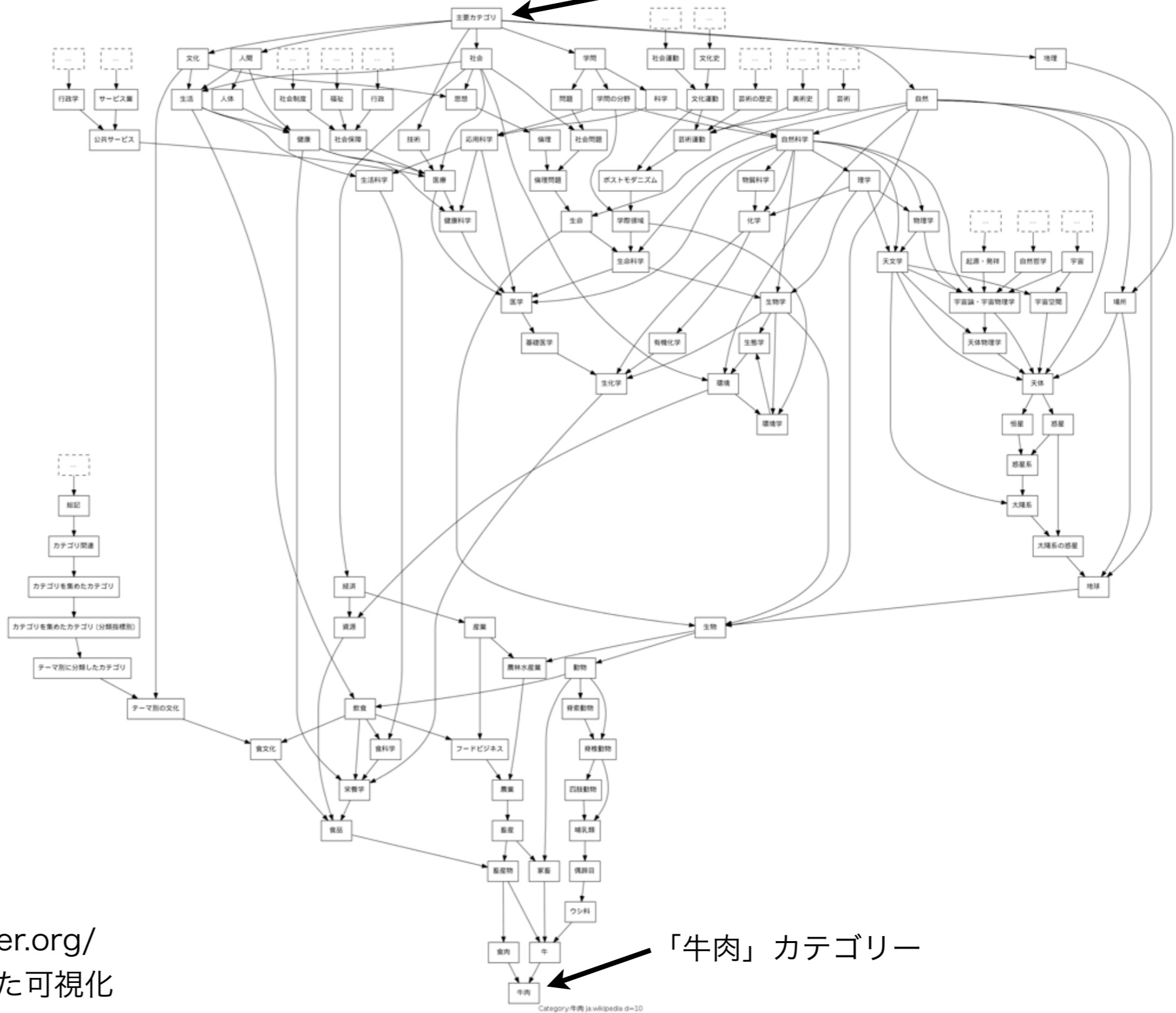
# モチベーション

- 文章がどんな Wikipedia カテゴリを持つか調べれば深い意味情報が分かる
- 質問応答の Question Classification などの応用が目的



# Wikipediaのカテゴリ

100個のルートノード



「牛肉」カテゴリにたどり着くまでのルート (一部)

「牛肉」カテゴリ

Catgraph (<http://toolserver.org/~dapete/catgraph/>) を用いた可視化

# モチベーション

- 1. Wikipedia のカテゴリ情報を使ってトピックをより厳密に取りたい
  - 教師ありトピックモデルが必要
- 2. 好きな粒度のトピックがほしい
  - 宇宙>惑星>木星
  - 階層的なトピック
- **教師ありの階層的トピックモデル**

# 階層化 LDA

- Hierarchical LDA (Blei+ 2010) (Blei+ 2003)
  - 教師なしで1層からL層までのトピックを構築・学習する
  - 今回はこのモデルを教師あり学習に拡張

