

デジタル技術によるアーカイブの開き方

～ “Japan Search”が目指す世界～

高野明彦

(国立情報学研究所／東京大学)

誤情報とファクトチェック

・”誤情報時代に我々はどうファクトチェックするか”

[R. W. Stevenson, New York Times, July 30, 2019]

- ファクトチェック機関: Politifact, Factcheck.org

・”最新のフェイク文章生成AIは危険過ぎて公開できない！ 作成者が公開見送り”

[Alex Hern, The Guardian, Feb 14, 2019]

- イーロン・マスクらが設立した OpenAI は、悪用を恐れて、研究成果の文書生成AIの公開を中止した

- GPT2

ファクトチェックのための4つの基本動作

[M. A. Caulfield, “Web Literacy for Student Fact-Checkers”]

- 先行例を探す: その主張について、誰かがすでにファクトチェックや検討を行なっていないか調べる
- 情報源を上流に辿る: 主張の根拠とされている論文や記事の情報源にあたり、信頼性を評価する
- 情報源の評判を調べる: 主張の根拠となった情報源について、他の人たちがどう評価しているか評判を調べる
- 最初に戻って繰り返す: 迷子になったり、自分が何をしているのか混乱したら、もう一度最初からやり直す

Carl Sagan on using space to sell Teslas

[Reddit 2018]



Carl Sagan with Pioneer Plaque



意思決定において陥りやすい罠

[John S. Hammond, et al.1998]

- Anchoring: 最初に見つけた情報から過度に影響を受ける
- Confirmation: 無意識に自分の既成概念を支持するデータを探し、それを覆す証拠は避ける
- Memorability: 直近の出来事や劇的な事件に過度に影響を受ける、複数の情報源から繰り返し同じ情報を受け取ると信用してしまう
- Status quo: 現状維持に役立つことを受入れ易い
- Sunk cost: 過去の過ちをなかなか認めずに、これまでの選択を正当化する方向で意思決定を行う

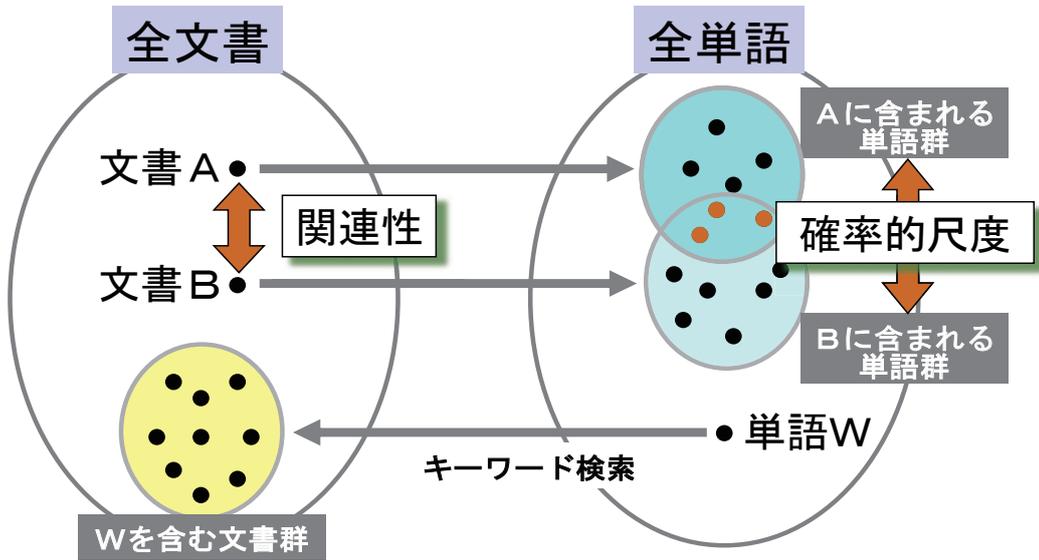
デジタル時代の罠に陥らない方法

- 概観: 関連する情報全体を広いコンテキストで把握し、それらの本質をつかむことができる
- 信頼性: 個々の情報の起源や由来について手がかりが得られる
- 要約: 集めた情報に共通する性質を自動的に取り出すことができる
- 飛躍: 注目している情報の周縁の情報を、観点を変えて眺められる
- やり直し: 情報を得たプロセスに縛られずに、無駄やロスを感じずに、情報探索を再スタートできる

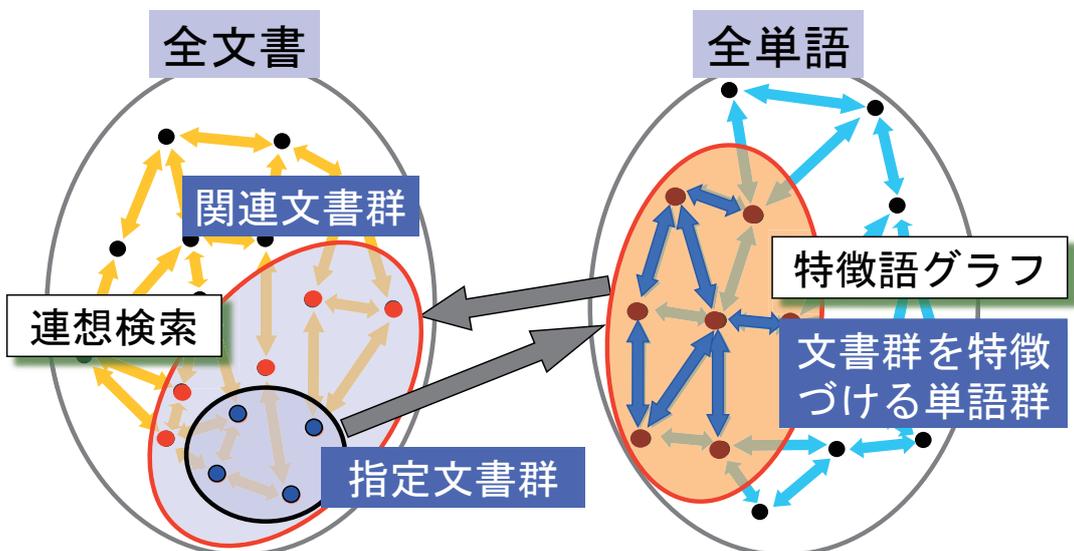
マニフェスト: 検索から連想へ

- リンク情報の高速処理によるキーワード検索 (リンク・クランピング)
→ 連想検索はコンテキスト計算 (コンテキスト・クランピング)
“情報のクオリティをひらめきに変える!”
- “水芸” から “蒸気機関” へ
ある場所に蓄えた情報を、内容を変えずに他の場所へ届ける
→ 蓄えられた情報を使って、他の情報を変換する
“コンテンツを変換する蒸気機関!”
- 人と情報空間の創造的相互作用
 - 人: 脳における記憶の連想的探索
 - 情報空間: 関連情報の探索・分析・提示**“情報空間は「第二の脳」となりうるか!”**

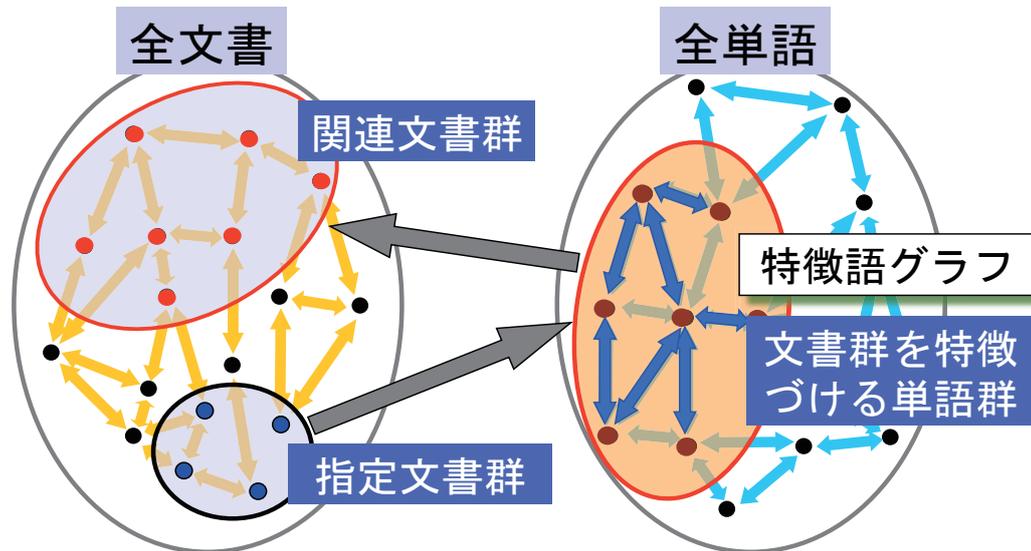
文書空間と単語空間の双対性



「関連性」を計量する連想計算



異なる情報源(サイロ)をつなぐ連想検索



“連想の情報学”

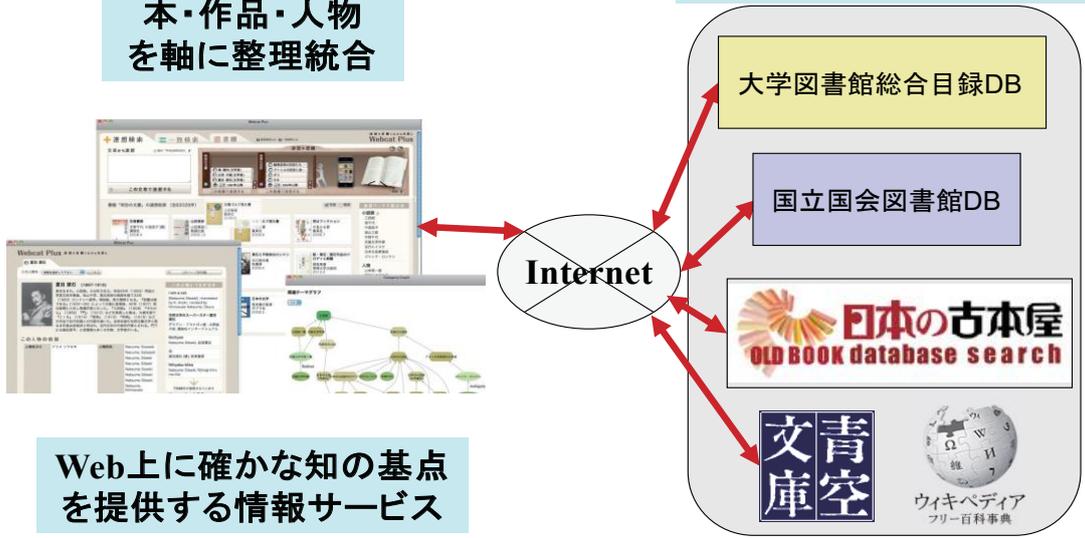
- GETA — 汎用連想計算エンジン
 - コンテンツを連想計算機構にコンパイル
 - 類似性により仮想的バックリンクを計算
 - 相互運用性のある連想計算Webサービス
- IMAGINE — 連想による情報源統合
 - 独立した連想の動的リンカー
 - 分散コンテンツ群の関連性フィードバック
 - 連想計算同士の相互作用

Webcat Plus (2010~)

＜連想×書棚で広がる本の世界＞

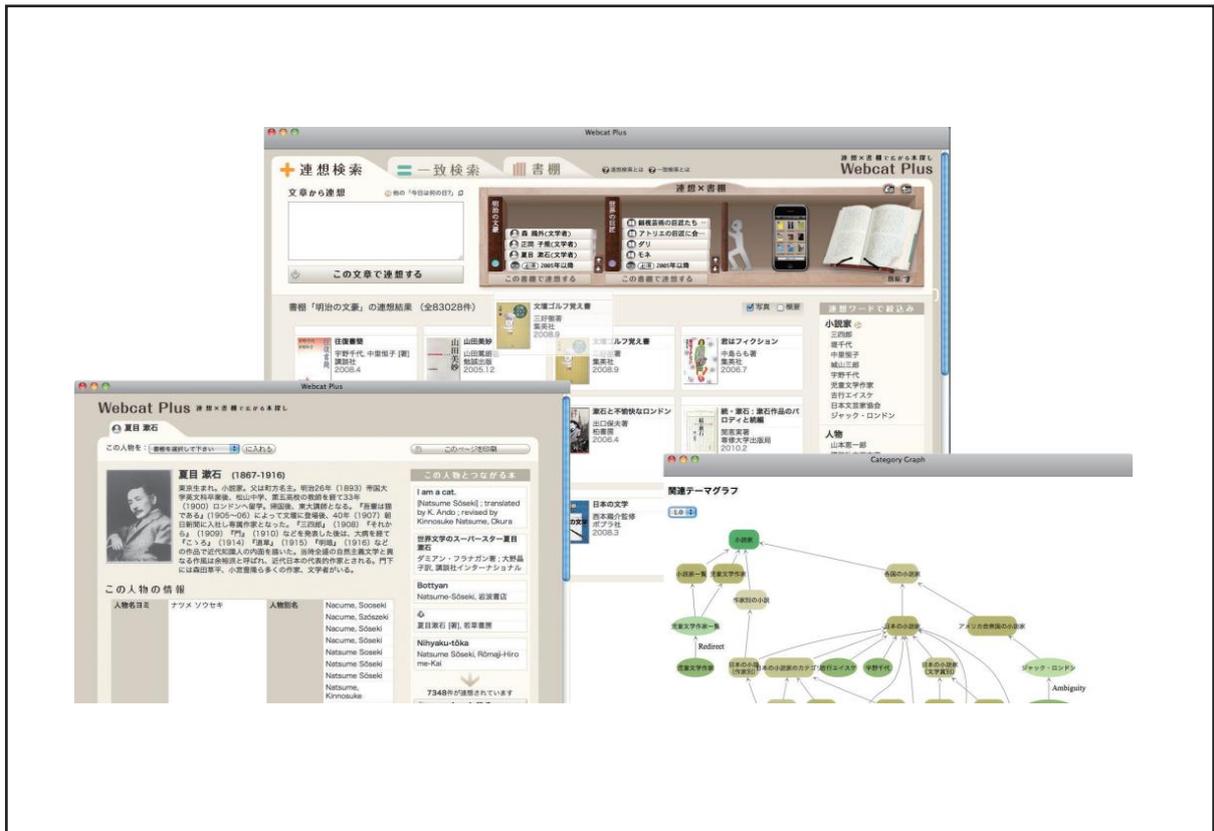
本・作品・人物
を軸に整理統合

各種情報源より情報収集



Web上に確かな知の基点
を提供する情報サービス





文化遺産オンライン(参加博物館・美術館:1000館、収録文化財:約26万件)

文化遺産オンライン <http://bunka.nii.ac.jp/>



電子企画展システム 遊歩館

