

タクソノミー、キーワード検索、 そしてフォークソノミー



平川正人

島根大学 総合理工学部/附属図書館

自己紹介

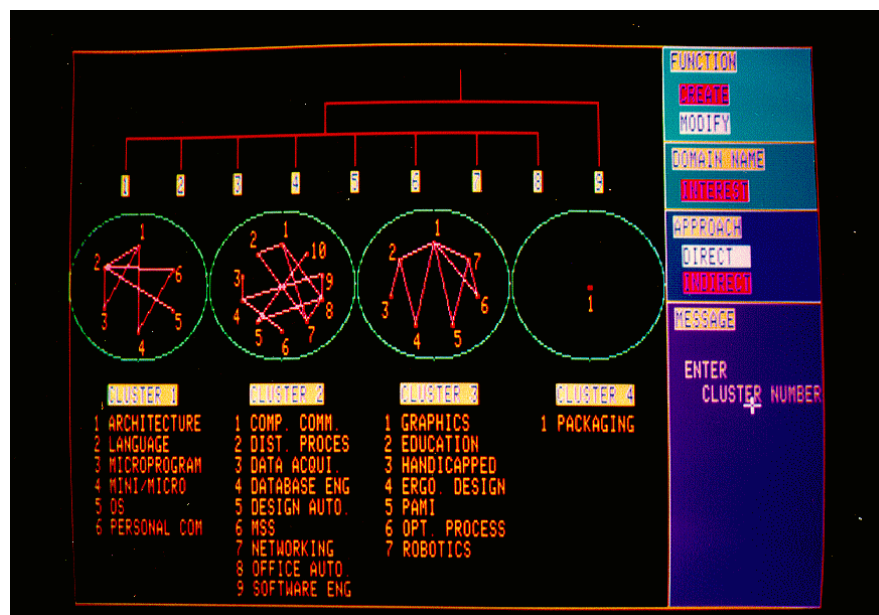
- あいまい検索システムARES (1980年頃)

MEMBERS	MEMBER_#	NAME	AGE	SEX
	_1000	P._ICHIKAWA		

PROFILE	MEMBER_#	FIRM_NAME	JOB_FUNCTION	T_INTEREST
	_1000	P._HIROSHIMA_U	P._PROFESSOR	P._A

CONDITIONS
_ADDATABASE_ENG.

JOB_FUNCTION	T_INTEREST
P._PROFESSOR	P._



タクソノミーに基づく情報検索

- Taxonomy is the practice and science of classification. The word comes from the Greek $\tau \acute{\alpha}\xi\iota\varsigma$, *taxis* (meaning 'order', 'arrangement') and $\nu \acute{o}\mu \omicron\varsigma$, *nomos* ('law' or 'science').
- タクソノミーは分類学と訳され、生物を分類することを目的とした生物学の一分野。生物を種々の特徴によって分類し、体系的にまとめ、生物多様性を理解する。なお、広義の分類学では、無生物も含めた事物(観念も含めて)を対象とする。歴史的には博物学にその起源があり、この世に存在する、あるいは存在したすべての生物をその対象とする。
- 図書館学、生物学、財務・会計、ソフトウェアなどの分野で広く採用されている。
 - 財務情報開示のための標準電子的ひな型(分類辞書):金融庁が主導するEDINETタクソノミ
 - 日本十進分類法(NDC)
 - 日本で使われている図書館資料の分類法で、社団法人日本図書館協会が作成。
 - 国立国会図書館分類表(NDLC)
 - 国立国会図書館の分類表で、1963年に初版、社会科学部門に重点を置く、アルファベットと整数の混合型非十進法。

タクソノミーに基づく情報検索

□ web検索での実現例(ディレクトリ型)



タクソノミーに基づく情報検索

□ 分類に基づくウェブサーチ(ディレクトリ型)の特徴

■ 利点

- 第三者の専門家が行なった木構造の分類に従って、ユーザは必要なサイトを探索する。
- 全体を概観しながら、用意されたシナリオ(階層構造)に従って検索を進めることができ、初心者にも使いやすい。

■ 弱点

- 分類語(カテゴリ)はいずれも人が作ったものであり、あらゆるサイトは、いずれか(ひとつ)の分類語に置かれる。
- それぞれのサイトは複数の話題を提供していることが多いが、分類を行なった専門家の思想を読み取れない限り、目的のサイトにはたどり着けない。

タクソノミーに基づく情報検索

- 国立国会図書館『日本十進分類法新訂9版』分類基準
 - 国立国会図書館における分類作業の基本方針と表の解釈を示す基準
 - 例) 548.2 電子計算機: デジタル計算機, アナログ計算機
 - 電子計算機工学<一般> は, ここに収める
 - 個々の電子計算機の入門書・操作法は, 548.29に収める
 - パソコン入門は, 548.29に収める
 - パソコン用語集(548.21/548.29に関するもの)は, ここに収める
 - 特定主題のパソコンの用語集は, その主題の下に収める
 - 例) 操作法に限定されている用語集 548.29
- “管理”することが主であり、利用者の視点は必ずしも汲み入れられていない。

キーワード検索

□ 検索エンジン(ロボット)

■ 利点

- 視点を強要されず、利用者が自由にウェブサイトを検索することができる。
- ロボットによるサイト情報の自動収集により、検索対象範囲が拡大。

■ 弱点

- 適切なキーワードを利用者が思いつかなければ使うことができない。
- 字面のみでのマッチングであり、無関係なサイトも提示される。

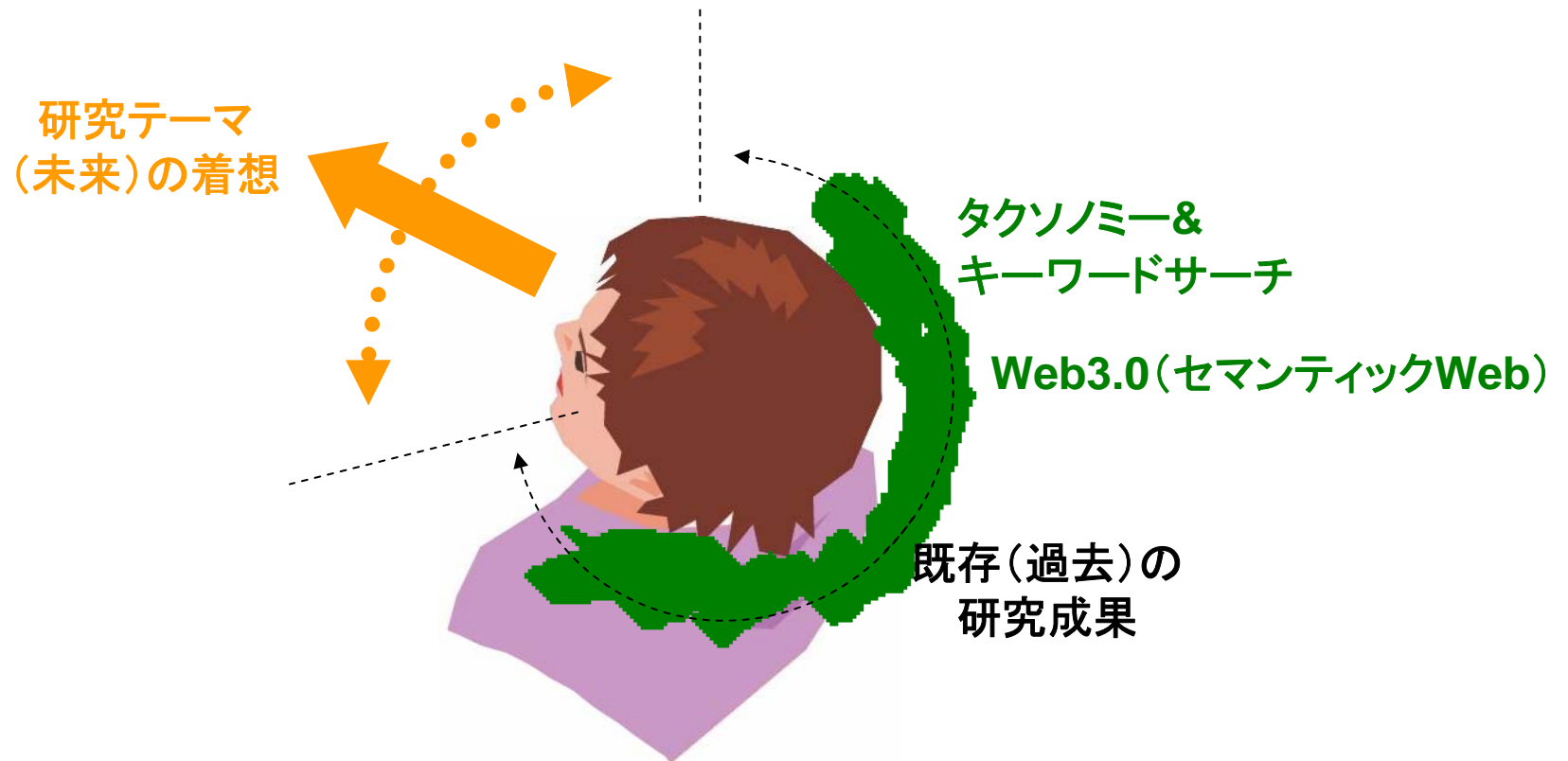


Web3.0

- シンタックス(構文)だけに頼るのではなく、セマンティックス(意味)を解釈することで、精度の高い検索を実現する。
→ Web3.0
- セマンティックWeb／Webオントロジー
 - コンテンツやデータの意味情報、すなわちメタデータを付加することで、コンピュータがコンテンツやデータをより高度な形で処理できるようにしている。
 - セマンティックWebはティム・バーナーズ・リー氏が1998年に提唱。

Web3.0

□ 大学教員(研究者)の視点から見た位置づけ



Web2.0

- オープンソース運動の推進役であるティム・オライリー氏が自身の論文『What is Web2.0』の中で紹介(2005年11月)。
 - 「明確な輪郭は持たず、他のものを引きつけるコアとして存在する」と定義
- Web資源の受け手が参加し、自ら情報を検索・受信・発信・共有をすることにより、スパイラル的に情報が蓄積されていくスタイル。
 - 例: ブログ、SNS、ソーシャル・ブックマーク
- ユーザーが自らの主観に基づいて好きなように属性(タグ)を与えることができる。しかも重要なことに、誰が、どのようなコンテンツに、どのようなタグを付けたのかが共有され、タグを介して同じ関心を持った他の利用者をつながることができる。
 - フォークソノミー
 - folks(民衆)とtaxonomy(分類法)を合体させた造語
- いわゆる「ロコミ」

Web2.0

- ソーシャルネットワーキングを取り込んだ新しい蔵書検索システムSOPAC (Social OPAC)
 - ミシガン州アナーバーの公共図書館AADL (アナーバー地域図書館: Ann Arbor District Library)が供用開始
 - 2008年9月にSOPAC 2.0が公開
 - オープンソース
- ファセット方式での検索結果ブラウジング、タグクラウド、利用者によるコメント・レビュー・レイティング情報などの付与機能。
- どのような統合図書館システム(ILS)とも連携してデータを取り込めるほか、利用者によるコメント等のデータを他機関のものと互換可能にしている。
- ILSシステムとの連携部分は“Locum”、利用者付与データとの連携部分は“Insurge”という名称で、SOPAC 2.0のエンドユーザ向けインタフェース部分とは独立。
- 類似のサービスとしてBiblioCommonsがある。

WebX.0 ?

