「信州サーチ」の未来設計

- これまで考えたこと、これからのこと -

吉本龍司 (カーリル) 第 4 回信州・知の連携フォーラム 2020/9/28





2

今日のテーマ「信州サーチ」 これまでに考えたこと - これからのこと

1

CC-0

自己紹介 カーリルのこと。





これまでに考えたこと。

5

7

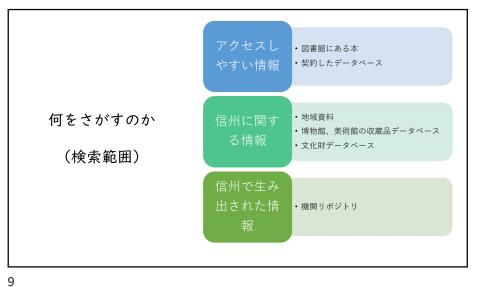
信州を検索できるようにしたい

- 図書の検索としての「信州ブックサーチ」
 - → 2016年にサービス開始(カーリルの検索技術を活用)
- 信州知のポータル
- 本だけじゃない検索
- 信州の文化、知識を検索する??

何がしたいのか

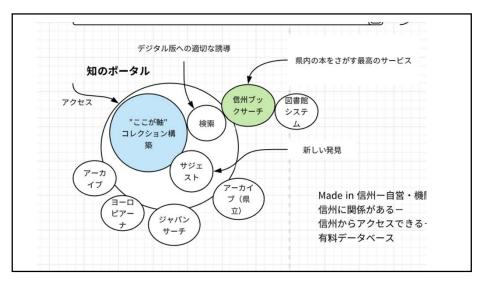
- 情報(知識)がある さがせない = あることを知らない
- あることがわかる -図書館のWeb-OPAC/収蔵品検索(コンテンツそのものの検索)
- ・簡単にあることがわかる -横断検索/総合目録
- アクセスできる -借りられる。見に行ける。
- 簡単にアクセスできる PDF、IIIF、デジタルアーカイブ
- 再生産できる -自由なライセンス

_



・図書館にある本 契約したデータベース やすい情報 何をさがすのか • 地域資料 ・ 博物館、美術館の収蔵品データベース る情報 文化財データベース (検索範囲) 機関リポジトリ 信州の機関以外がもっているかも Exp.ジャパンサーチやヨーロピアーナ

10



検索するために、 塊(コレクション)をつくる

検索技術の動向

Google誕生から22年



1998年9月27日、インターネット検索サイトの「Google」が誕生しました:今日は何の日? https://japanese.engadget.com/today-203006427.html

集中型と統合型

検索するためには、集めなければならない 集めるためにはコストがかかる

13

- 検索サービスをひとつに統合したい―横断検索 おそい、安定して動作しない、同じデータが重複する
- カーリルでは、これらの議論を現実的手法で解決 (少なくとも図書館の分野では…)

ワンボックス時代はとっくに終わってる

- ・いわゆるGoogleワンボックスではない 地図の検索、図書の検索、レシピの検索 日々、たくさんの検索を使っている
- •目的やニーズに合わせたスコアリング 検索は恣意的
 - 売れそうな本を上位に

14

- 役に立ちそうな本を上位に
- 信頼できる情報を上位に
- 満足度の高いレシピを上位に
- 違った目的をもった情報を集めてもなかなかうまく機能しない









検索の民主化

- かつて検索は極めて高コスト
- あいまい検索、日本語形態素解析、転置インデックス
- 多くの基礎技術が共有されるようになった
- ・数兆件規模の検索であっても、オープンソースで運用できる
- だったらもっと自由にできるはず
- ・見直される横断検索に関する技術 (metasearch-engine)



Exp. 検索エンジンの横断検索 複数の検索エンジンを統合的に検索できる =誰が何を消しているのかがわかる これからのこと。

21 22

どんなサービスにしていくか

- もちろん、カーリルも積極的に関わっていきたい
- 長期的なミッション
- •現実的な取り組み

何が検索できるべきなのか

- 今の技術やシステムでできることにこだわらない
- そのためにどのような技術が必要か
- ・課題は何か

連携する機関にとってのメリット

- ・カーリルは、全国7300館以上の図書館と連携
- ・最初は… 仕事が増えて困るとの声も
- ・館内の検索端末で、ほかの図書館の本も探せる!そのままリクエストもできますよ!
- •信州サーチとつながったらどんなメリットがある?

無断で使われることを増やす

- 博物館職員が文化財情報の利用を制限する前に考えておくべき リスク
- •公開して、検索できるようにしておくだけ
- 本当に許諾はいるか?

25 26



高校生のときに作ったウェブサイト



27

ウェブサイトに掲載することは 出版と同じではないか

- だとしたら図書館は当然に探せるようにする必要があるし
- 利用についていちいち断る必要もない
- ・社会的合意の形成、MLA機関が一歩先へ 柔軟なライセンスの適用 etc クリエイティブコモンズ "止める権利のないものを止めない"

翻訳技術の進歩

- 信州に関する情報は日本だけにあるわけない
- もう僕たちは、お互いに違う言語環境でも、 自由に知識を共有できる
- DeepL

29 30

(勝手に) ミッションを考える

検索のニーズをひろげる

新しい知識を生み出す

コラボレーションのスピードを加速する

33



あたりまえのインフラとして