

検索エンジンの歴史的意義 と未来

森大二郎

未来検索ブラジル

検索エンジンの・・・

- 歴史的な位置づけ
- 現状と課題
- 未来

について話します

検索エンジン
とは
何なのか

検索エンジン
は、人類のコ
ミュニケーション
の質的変化を
促す道具である

歴史的に
考えてみる

人類史上、コ
ミュニケーション
の質的変化は
過去に2度あった

今回しよ

3度目!

1 度 目

言

語

2 度 目

文

字

3 度 目

検査工

ニニニ

-
- 1回目： 言語
 - 2回目： 文字
 - 3回目： 検索エンジン

ち

よ

こ

と

待

て

も っ と あ

る だ る

本，印刷術，活版印
刷，写真，電信，無線
通信，蓄音機，電話，
映画，磁気記録，TV，
写真，デジタル組
版.etc..

それらも確
かにスゴい
んですけど

ま、一じ
の意見で
すから

じやあ、な
ぜこの3つ
なのか

1 度目： 言語

- 音声言語を中心に発達
- 抽象的な事物, 複雑な関係の表現能力を飛躍的に高めた
- 時間と空間を共有する人との情報伝達
- 主に1:1のコミュニケーション

2度目：文字

- 時間と空間の制約を越えた情報伝達
- 1:多のコミュニケーションをスケール可能にした
↑ここ大事

その後

の

本，印刷術，活版印
刷，写真，電信，無線
通信，蓄音機，電話，
映画，磁気記録，TV，
写真，デジタル組
版.etc..

これらは
量的な変
化

1: 多々の情報
流通量を増
幅させた

検索エンジン
シミュレーションは？

3度目： 検索エンジン

- 情報の取捨選択(フィルタリング)を支援
人間の情報処理能力の補完
- 多:1のコミュニケーションをスケール可能に
- 検索エンジン=諸技術の総体と受け止めてね
 - 計算機, 電子化文書, ネットワーク, WWW, データベース, 自然言語処理, AI, 機械学習, 分散処理etc..

整理

- 言語

1:1のコミュニケーション

- 文字(その他情報蓄積・情報通信技術)

1:多のコミュニケーション

- 検索エンジン

多:1のコミュニケーション

検索エンジン
はまだまだ
たばか

過去の二つの
質的変化はど
のように進ん
だか

言語

- ある時突然誰かが発明したようなものではない
- 人の言語能力の多くは生得的な能力として組み込まれている
- 数万年という長い進化的な過程を経て発達した

文字

- 人類史上数回独立に発明された
- 絵 → 象形文字 → 表音文字 のように発達
- 音素文字(アルファベット)
 - 汎用性, 利便性において画期的な技術革新
 - 人類史上ただ一度だけ発明された
(by セム系労働者 @BC20c頃)
 - そこから派生・参考にされ世界中に普及
- 音素文字に至るまで千年の歳月を要した

検索エンジン

- 非常に歴史が浅い

- クランフィールド実験(1959)

- WWW (1991)

- Yahoo (1994)

- Google (1998)

- 多:1のコミュニケーション手段としての検索エンジンが発達するには、エンジンとユーザの両者に変革を必要とする

現狀と

課題

課題なんてあるの？ Google
で十分じゃないの？

ま

す

い

違

検索エンジンとは

- 知りたかった情報に
- 素早く
- たどり着くための道具

「知りた
かっただ情
報」とは？

そもそも

「情報」

とは？

情報とは何か

- 「分からなさを減らすもの」
- 情報理論的定義
(by C. シャノン)

情報とは何か(2)

- 「差異を生み出すための差異」
(by G. ベイトソン)
- 「差異」が2回出てくる
- 二つ目の「差異」は
 - 一様・ランダムでない「パターン」全般を指す
 - 全てのパターンは情報になり得る
 - 一方、本に書かれた文字であっても情報と言えない場合もある

情報とは何か(3)

- 「差異を生み出すための差異」
- 一つ目の「差異」は
 - 何らかの意味を構成することを指している
 - 受容・解釈されて「パターン」は「情報」になる
 - 解釈と情報は不可分

つまり

- 何が情報で何がノイズかはユーザによって異なる
- 検索文字列に対応する正解集合はユーザによって異なる
- 「知りたかった情報」とはそういうもの
- 現状の検索エンジンはこれに対する解を持たない

でも・・・Google
使ってって特にな
問題を感じな
いけど？

歴史を振り返る

Google以前(1994~1998)

- 文書の内容そのものからスコアを算出
- tf・idf法とか
- 文書の内容に作為的な操作がなければ有用
- しかし、SEO耐性が低い
 - SEOスパムが広まると急速に精度が低下した

SEO

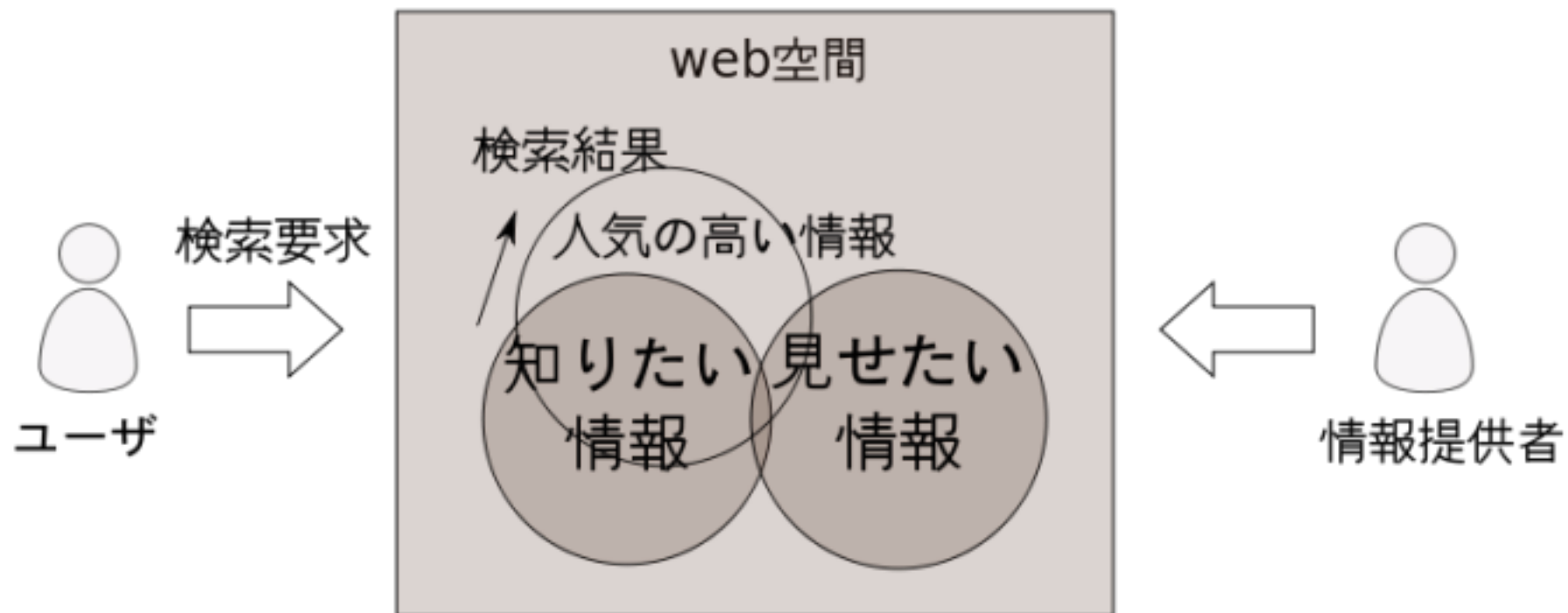
- 検索結果を情報提供者の都合に引き寄せようとする行為



Google以降(1998～)

- ページ毎の人気の高さからスコアを算出
- 「人気の高い情報」 ≡ 「満足度の高い情報」
- 人気の高さ(=第三者の評価)は、文書の内容と比べて情報提供者にとって操作しにくい
- →SEO耐性が高い

人気の高い情報



しかし

- 「知りたかった情報」 ≠ 「人気の高い情報」
- 人は知りたかった情報が不確かな段階で検索を開始する
- 「人気の高い情報」に出会った時点で「満足」してしまう
- 「満足度の高い情報」が有用とは限らない
- → 知りたかった情報に出会わずに終わる

「満足度」はトラップ

- 商用検索エンジンの目的はトラフィックを集め、メディア力を高めること
- 「満足度」が高ければユーザはリピートする
- エンジンとユーザの利害が一致し、そこで安定してしまう
- エンジンとユーザの両者にとって共通のトラップ
- 一度嵌ると抜け出すのは困難

検索エンジンの未来

- 多:1コミュニケーションの手段として成熟
- 「知りたかった」情報との出会いを促す

出会いを促す

- 文字の歴史の話に戻りますが..
- 音素文字は、多くの人々の交流と出会いによって生み出された
- 革新的な発明は、一人の天才の力だけでなく、人やアイデアの出会いによって生み出される
- 出会いを促進し、人の持つ能力を高い次元に押し上げるのが検索エンジンの責務

どうやって実現するのか

- まだわかりません。そう簡単でもないでしょう。
- エンジンとユーザ両者が成熟するのに長い時間が必要かも
- 100億のページをクエリに適合する順に1次元に整列させるのは限界があるかも
- Sennaには僕らなりのアイデアを実現するための機能を組み込んでいます
- なまあたたかく見守ってください

ご静聴ありがとうございました
がとらうござ
いました

質疑応答

Q: Googleのスコアは「人気の高さ」だけではないでしょ？

A: もちろん様々な要素を加味していると思うが、
コンテンツベースのスコアに比べてリンクベースのスコアを重視したのが要点だったと思う。

Q: ユーザ毎の正解の差異は、パーソナライズド検索技術によって解決されるのでは？

A: パーソナライズド検索は重要な要素技術。
しかし、現状の技術は、人を類型化して扱うのが基本原理。それだけではダメだと思う。

「あなたがどんな人間かは、
あなたがどんな音楽を聴いていて、
どんな服装をしているかでわかる。
でも、そんな分かれ方って屈辱だろ？」 (by 橘川幸夫)