開発運営を続ける楽しさ
ニュース検索サイト CEEK.JP NEWS の開発を通じて

吉田光男
筑波大学大学院システム情報工学研究科
有限会社てっくてっく / CEEK.JP

静岡大学 大学院GP

2010/7/5

アウトライン

Ceek.jp

- 自己紹介
- Web検索エンジンの基礎・課題
- 事例
  - CEEK.JP
  - CEEK.JP NEWS
  - 有限会社てっくてっく
- キャリアを考える
- おわりに

自己紹介

Ceek.jp

- 吉田光男（よしだみつお）
- 所属
  - システム情報工学研究科 CS専攻
    - 自然言語処理 on the Web 研究室
  - 産学リエゾン共同研究センター 客員研究員
  - 有限会社てっくてっく
- 1984年生（25歳）
- 和歌山県出身

経歴

Ceek.jp

- 2003年 情報学類 入学
  - AC入試（AO入試）
    - 図書館の情報をインターネットに解放するぞ！
- 2005年 留年（2年生2回目）
  - 年度末に起業
- 2006年 留年（2年生3回目）
- 2009年 システム情報工学研究科 入学
  - 研究漬け
- 2011年 進学できたらいいな…

所属（1/3）

Ceek.jp

- システム情報工学研究科 CS専攻
  - 自然言語処理 on the Web 研究室

ACCC Photo Archives, Univ. of Tsukuba
http://photo.cc.tsukuba.ac.jp/

所属（2/3）

Ceek.jp

- 産学リエゾン共同研究センター
  - 客員研究員

ACCC Photo Archives, Univ. of Tsukuba
http://photo.cc.tsukuba.ac.jp/

Shizuku Lab.

株式会社しずくラボ

小野永貴
図書館情報メディア研究科

所属（3/3）

Ceek.jp

- 有限会社てっくてっく
  - 代表取締役

・サービス運営
・受託開発
・コンサルティング

Web検索エンジンの基礎

拡大するウェブ

Ceek.jp

- 大量のウェブ
  - 2,600万 unique URLs (1998)
  - 1兆 unique URLs (2008.07)

We knew the web was big…
(Official Google Blog, 2008)

- 拡張するウェブ
  - 企業サイト
  - eコマース
  - SNS（ブログ）
  - 画像, 動画

tsukuba.ac.jp


amazon.co.jp

twitter.com


youtube.com

ウェブの道しるべ

Ceek.jp

- 必要な情報を得るために
  - ウェブ検索エンジン
- ウェブ検索エンジンの利用率
  - 93.7% （日本 2008）

インターネット検索エンジンの現状と市場規模等に関する調査研究
（総務省情報通信政策研究所, 2009）

- 主要なウェブ検索エンジン
  - Yahoo!
  - Google
  - Bing （マイクロソフト）

yahoo.co.jp

google.co.jp

bing.com

検索エンジンの分類

Ceek.jp

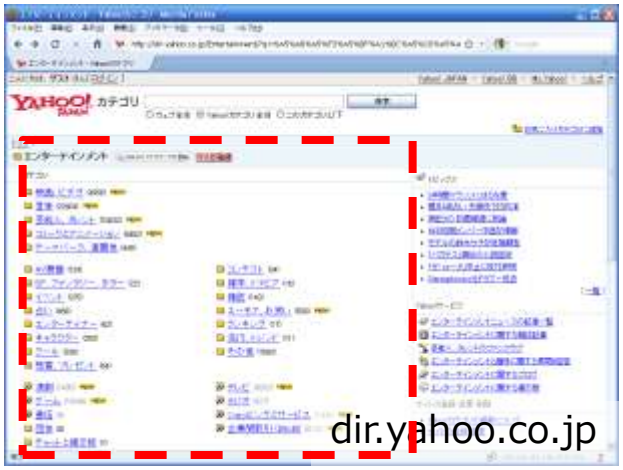
- ディレクトリ型（Yahoo!）
  - サーファーが収集（人力）
  - 信頼性の高いウェブサイト
  - 登録サイト数に限界

フルテキスト型

- ロボット型（Google）
  - クローラが収集（自動）
  - 大量のウェブサイト
  - 質の低いサイトが混ざる

本当の検索エンジンの分類

Ceek.jp



人力

ディレクトリ

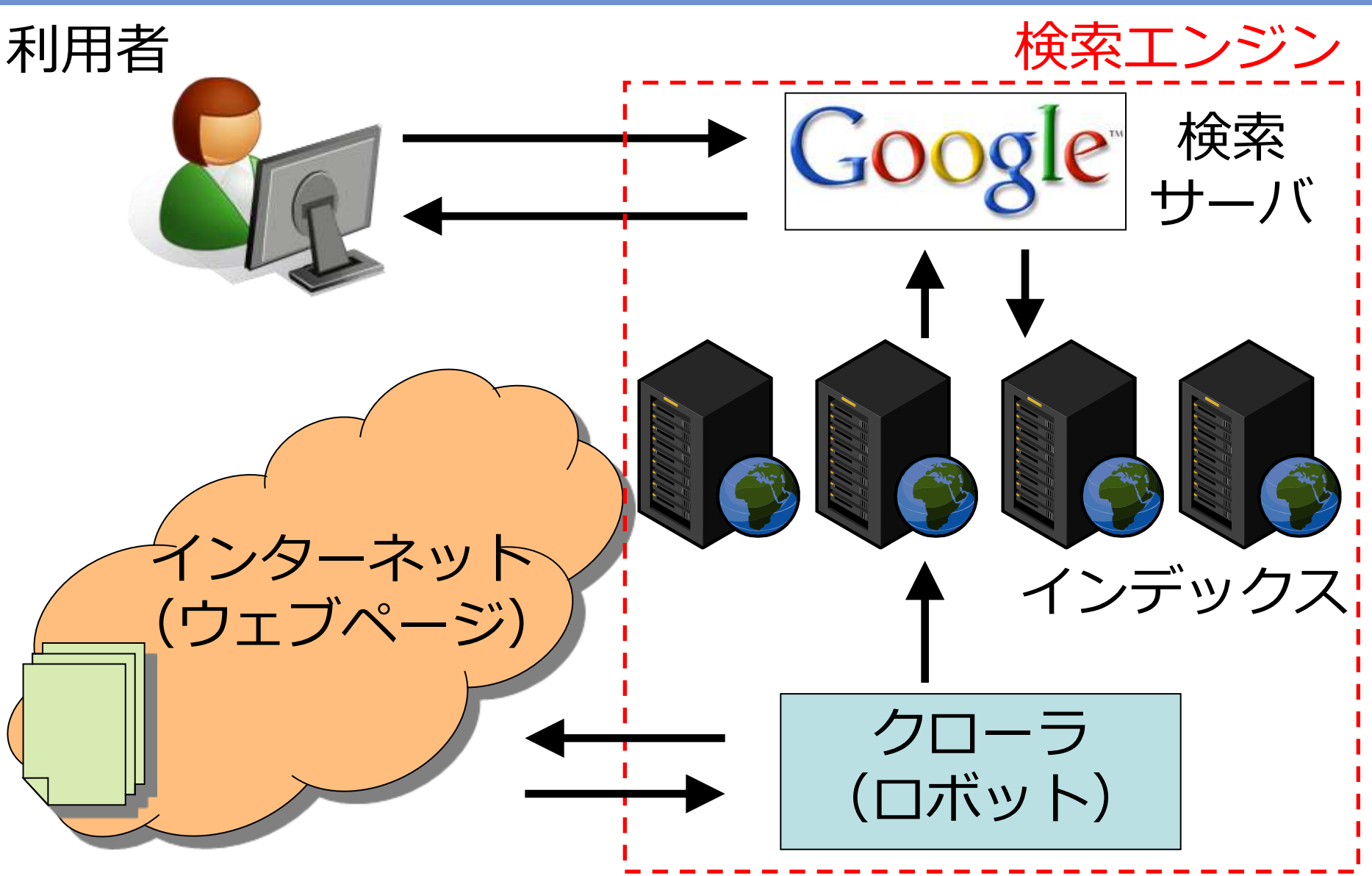
フルテキスト

ロボット

ここをどうするか？

ロボット型検索エンジンの全体像

Ceek.jp



Web検索エンジンの課題
未来の検索エンジン

キーワードや求めているものがはっきり分かっていなくても、探し出す検索の仕組みを作りたい。それが本来の検索のはず。

（『日経産業新聞』2010年5月18日 6面）

キーワード検索の限界

Ceek.jp

- 内容との乖離

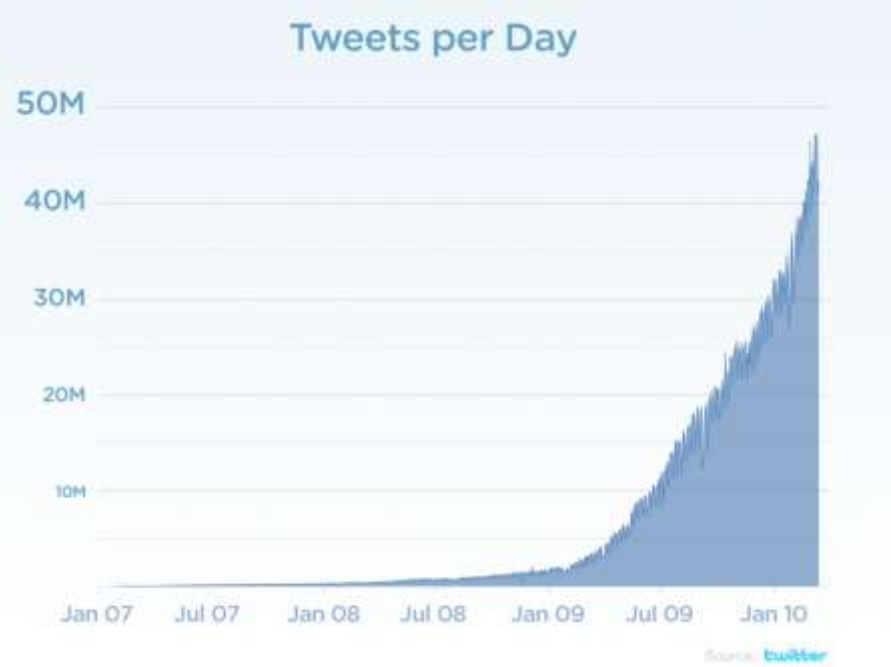
可愛い犬ですね！

- 概念で検索
  - あるキーワードが含まれるウェブページ
  - 「元気の出るウェブページを見たい」

ページ単位の限界

Ceek.jp

- マイクロブログの普及
  - ショート・コンテンツ



ツイートの測定結果
（Twitterブログ, 2010）

前後の文脈がわからない

新しいコンテンツのランキング

Ceek.jp

- 大量の検索結果を全部みる？
  - 平均 2.35 （米国 1999）

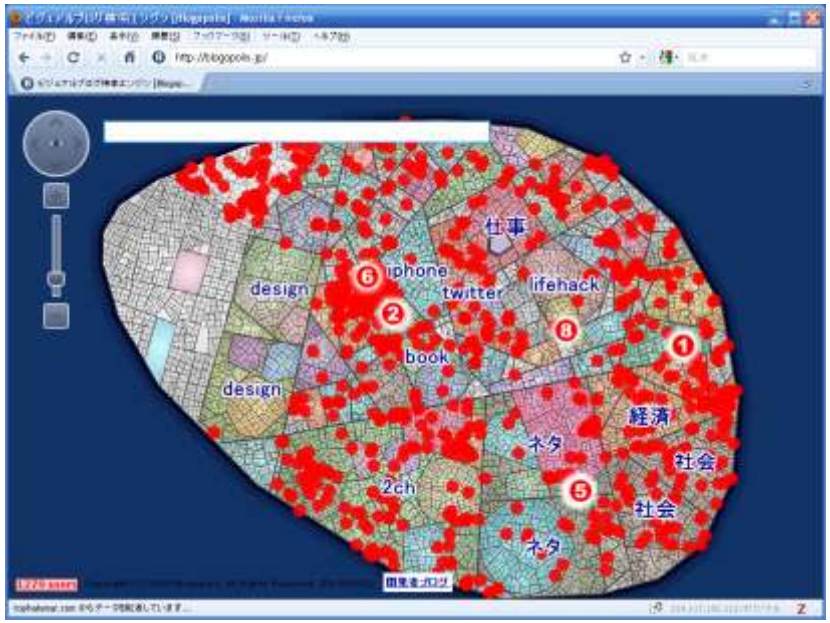
Amanda Spink, Jack L. Xu.
Selected results from a large study of Web searching: the Excite study.
Information Research, Vol.6, No.1, 2000.

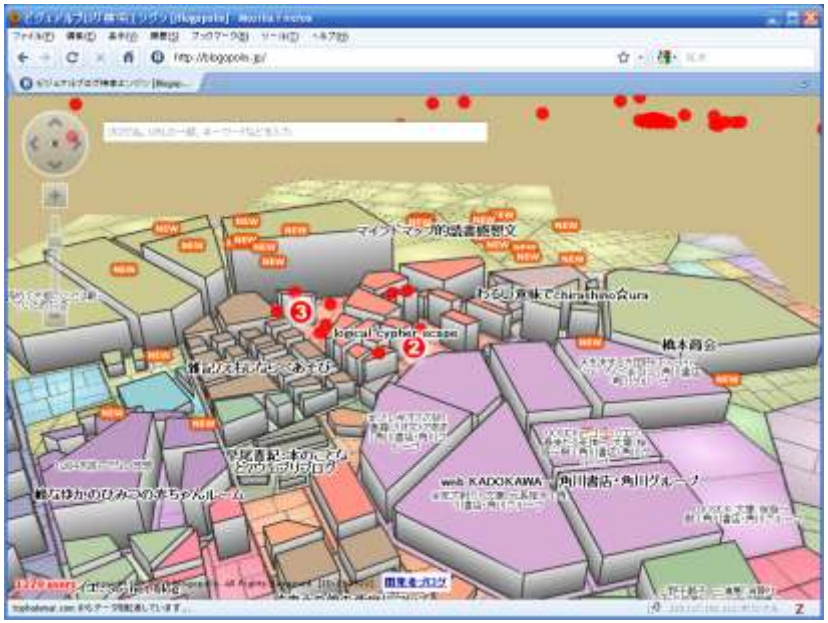
- ランキングアルゴリズム
  - ベクタスペースモデル, PageRank
- 新しいコンテンツ
  - ブログ, ニュース
  - ショッピング

検索結果の提示方法

Ceek.jp

- 電話帳メタファからの脱却
  - Blogopolis
  - http://blogopolis.jp/





CEEK.JP

統合型メタサーチエンジン

「検索」の技術が躍進してこそ、人類の英知の利用が促され、より良い社会になる。

（CEEK.JP ヘルプページ）

CEEK.JP（1/3）

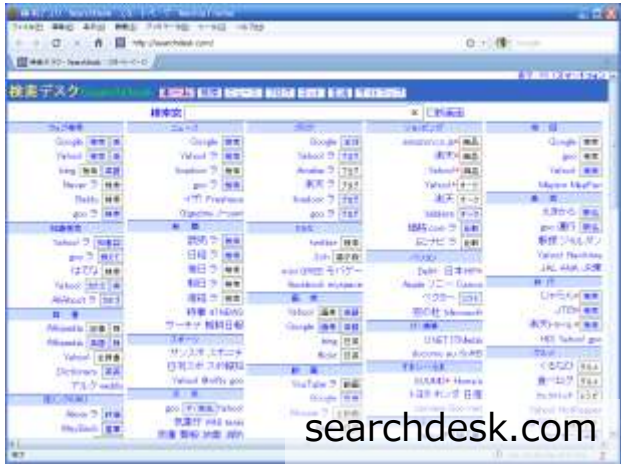
- CEEK.JP
  - 統合型メタサーチエンジン
  - 2002年8月開始

- 当時のウェブ検索エンジン
  - 定番が無い状態
  - 複数の検索エンジンを使い分ける

CEEK.JP（2/3）

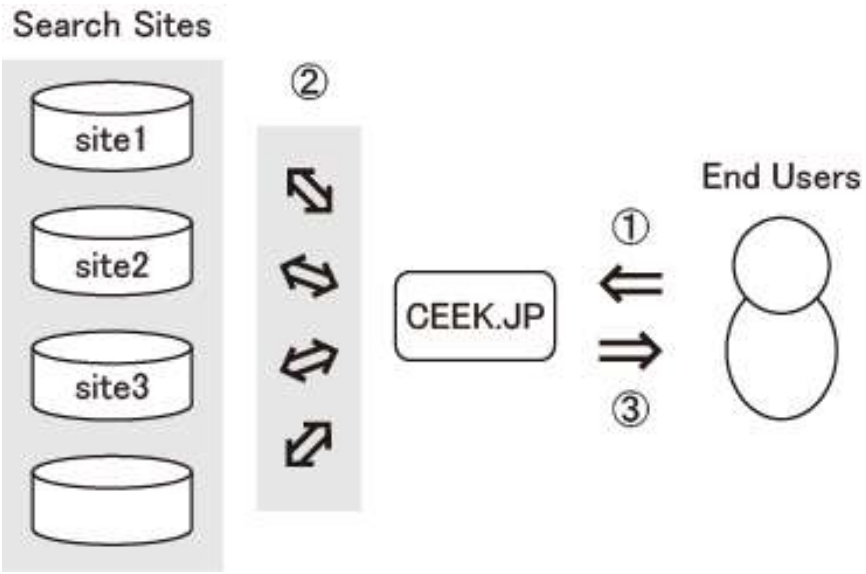
Ceek.jp

- 当時の検索方法
  - 検索サイトAで検索
  - 検索サイトBで検索
  - 検索サイトCで検索
  - 検索サイトDで検索 …



- CEEK.JPのねらい
  - 作業の自動化
  - 検索サイトA～Dの結果を1ページにまとめる

CEEK.JP（3/3）



1. ユーザがキーワードを入力
2. CEEK.JP が複数の検索エンジンで検索
3. 検索結果を統合して表示

CEEK.JP NEWS

ロボット型ニュース検索エンジン

CEEK.JP NEWS（1/6）

- CEEK.JP NEWS
  - ロボット型ニュース検索エンジン
  - 2004年6月開始（β版: 2003年11月）
  - 日本語では最も歴史がある

- 初期（2003年2月）
  - CEEK.JPと同じ仕組みで
    - ニュースソースの少なさ
    - 反映の遅さ
    - 権利問題

CEEK.JP NEWS（2/6）

Ceek.jp

- 需要の予想
  - 海外でBlogがブームに
  - ニュース検索の需要が伸びると予想
  - 狭義のBlog（Wikipedia）
    - ウェブページのURLとともに覚え書きや論評などを加えログ（記録）しているウェブサイトの一種
- 当時の（僕の）状況
  - 日本語のロボット型ニュース検索が無かった
  - ウェブ検索エンジンを作りたい

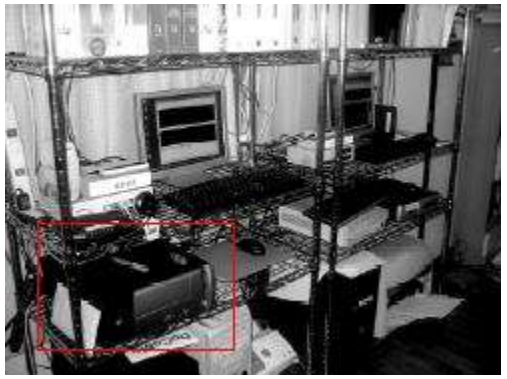
CEEK.JP NEWS（3/6）

Ceek.jp

- サーバを準備しないと
  - 共有サーバでの運用は難しい（特にクローラ）
  - 専用サーバは高い
  - 自前でサーバを構築しよう！

「ママ、どうしておうちにサーバーがあるの？」
マイクロソフト社が作成したパンフレットのタイトル

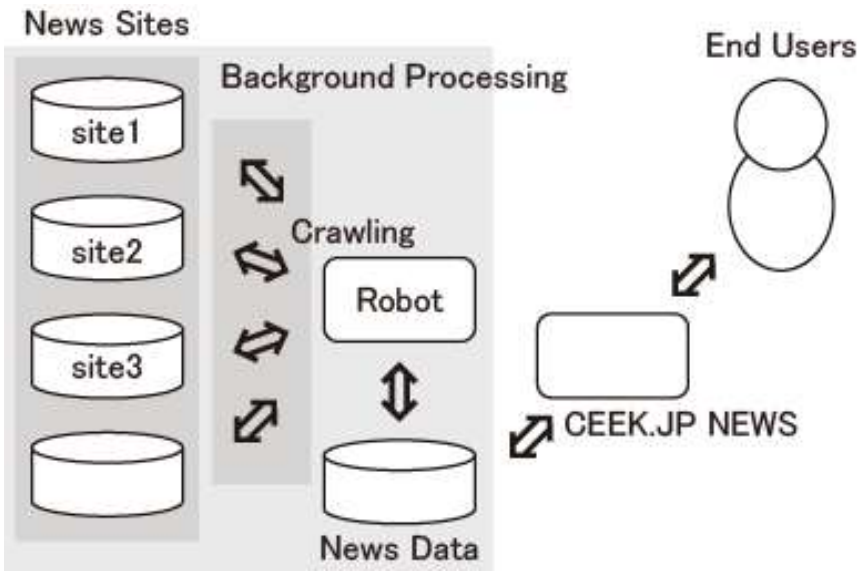
- メリット
  - 暖房要らない :)
  - ハードウェアの自由度が高い（メモリ, SSD）



初期（登宅）

現在（自宅）

CEEK.JP NEWS（4/6）



- 前述のロボット型検索エンジンと同様
  - リンクは1ページのみ辿る
  - サイトに応じた解析エンジン
    （タイトルや記事本文の抽出）

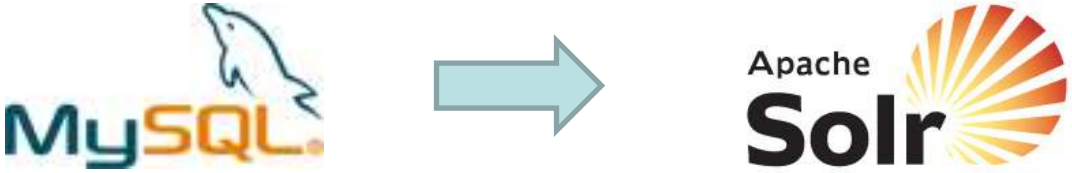
CEEK.JP NEWS（5/6）

Ceek.jp

- サービス理念
  - システムは情報の取捨選択をしない（表示順序は日付順）
  - 機械化が進むと人間は馬鹿になる（と僕は信じている）

- 検索エンジンのパワー
  - 興味がわかる
  - 未来の予測

- 今後の展望
  - 対象サイトの増強
  - バックエンドシステムの変更
    （ファセット検索の実現）



  - 有料検索サービスの提供
    - 解析結果の提供
    - API の提供
  - 未来年表の自動作成

有限会社てっくてっく

大学に入ってまっ先に思ったのは、人材がもったいないということ。
（『日経クリック』2007年2月8日 p.21）

てっくてっく（1/5）

Ceek.jp

- 有限会社てっくてっく
  - 2006年3月 設立
  - 変な会社名…
- 目的
  - 学生に適切な報酬を
  - サービス運用の法的リスク軽減
  - 新しい働き方を目指して（あとで）
- 有限会社
  - 決算の公開義務がない

てっくてっく（2/5）

Ceek.jp

- 学生に適切な報酬を
  - 学生の技術力は低いか？
  - 受注金額の50%を報酬に
    （ただし、継続雇用を保障しない）

- 失敗…
  - 高い報酬を避ける（責任回避？）
  - 自己マネジメントを行わない
    - 安定雇用を望む

てっくてっく（3/5）

Ceek.jp

- サービス運用の法的リスク軽減
  - 個人: 無限責任 / 法人: 有限責任

- （旧）著作権法
  - ウェブ検索エンジンの法的リスク
    - クローリング（複製権）
    - インデキシング（翻案権、同一性保持権）
    - 検索結果表示（公衆送信権）
  - 黙示の許諾論
    - 回避手段を取らない場合は許諾したと見なす

てっくてっく（4/5）

Ceek.jp

- （改正）著作権法（2010年1月1日施行）
  - ウェブ検索エンジンの例外追加
  - 第47条の6、政令第299号、文科省令第38号
- ライントピックス事件（平成17（ネ）10049）
  - 読売新聞が記事見出しの無断複製により著作権を侵害されたと提訴
  - 知財高裁は記事見出しの著作権を認めなかったものの、不法行為を認め賠償命令

てっくてっく（5/5）

Ceek.jp

- ビジネスモデル
  - 受託開発
    - 運営サイトを見た会社が依頼
    - ウェブサービス全般
  - 保守業務
    - 会社運営コスト（人件費等）の70%をカバー

- おかげさまで
  - 初年度からずっと黒字

キャリアを考える

一例です。自身のキャリアは自己責任です。

会社設立まで（1/2）

Ceek.jp

- PCに触る
  - 1995年（小学5年）
- ソフトウェア検索エンジン（K-Search）
  - 2000年（高校1年）
- 無料Webスペースの提供（K-Server）
  - 2001年（高校2年）
- CEEK.JP 開始
  - 2002年（高校3年）

会社設立まで（2/2）

Ceek.jp

- フリーランスで活動開始
  - 2003年（大学1年）
- アルバイトしてみる
  - 2004年（大学2年）
- フリーランスで凄く儲かった
  - 2005年（大学2年2回目）
  - 価格.com の類似サイト（提供終了）
- 会社創業（2006年）
  - 池嶋俊（現在はGoogle）と2人で

会社の中で

Ceek.jp

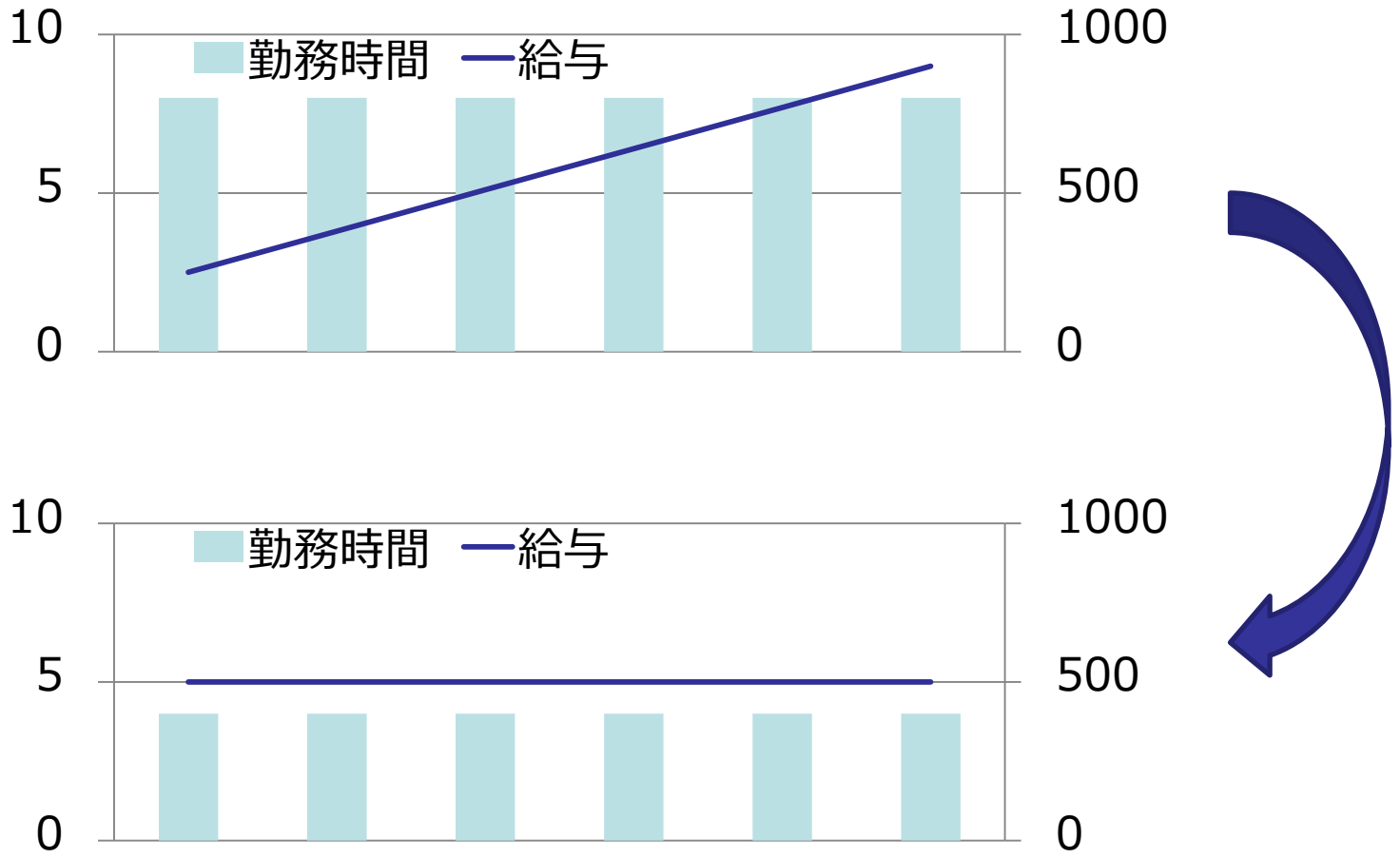
- 受託開発の方向性
  - ウェブとネイティブソフトのハイブリッド
  - 両方出来る会社はそんなに無い
    - 吉田: ウェブ
    - 池嶋: ネイティブソフト

ONPOO.net
柔軟な視聴設定ができ、P2Pで音楽を配信するサービス

時間と給与（1/2）

Ceek.jp

- どのように仕事をしたいか



時間と給与（2/2）

Ceek.jp

- 通常の給与モデル
  - 勤務時間は一定で給与が上がる
  - 勤務時間に選択肢がほぼ無い
- 目指したいモデル
  - 勤務時間を短くしてほどほどの給与
  - 自由な時間を増やす
    → お金を稼ぐために使う時間を減らす
- 起業するしかなさそうですね…

ほどほどを達成するために

Ceek.jp

- 5年で○億貯める！
  - 5年後には隠居生活
  - 身体壊しそう…
  - （かつてのベンチャー企業）
- ほどよくほどほど
  - 保守業務
    - 年間保守料の目安: 受託金額の30%
    - 品質の高いソフトウェアを納品しましょう
  - ウェブサービスの提供
    - 課金、広告収入

ちなみに…

Ceek.jp

- お金はあまり使わない
  - 食事
  - 図書
  - サーバ

研究と仕事

Ceek.jp

- 大学院生なので研究します
- 研究と仕事は両輪でありたい
  - 仕事上の課題を研究で解決する
  - 学術界との課題にずれが出るが…
    - 工学のゴールはビジネスなのでは？
- Webコンテンツ抽出
  - ニュース検索の運営コスト軽減

Webコンテンツ抽出

Ceek.jp

やり過ぎ注意

Ceek.jp

- 2009年度の対外発表
  - 論文誌1件、口頭5件、ポスター4件
  - 燃え尽きた…
- やりたい事と違うような…
  - 研究のための研究に
  - 本当にやりたい事は？
    - ユーザにプロダクトを使って貰いたい！

ウェブサービスの開発

Ceek.jp

- 手軽に開発・手軽に利用
  - ネイティブアプリにはダウンロードの障壁
- 上流から下流まで
  - アイデア、設計、開発、宣伝、運用 …
  - 手応えを感じる事ができる
- フィードバック得やすい
  - アクセス解析

サービスを開発していて

Ceek.jp

- 難しいのは「継続」すること
  - サービスの開発は比較的簡単
  - 継続する事が難しい
- 継続を容易にする2点
  - 自身が使う事
  - メンテナンス・フリーである事

継続する事

Ceek.jp

- 継続する事は自身の名を売る事
  - 「静岡大の○○です」以外は言えますか？
  - 続けていれば誰かの記憶にひっかかる
  - 継続は信用の基
- その結果
  - お仕事が来たり
  - インタビューされたり
  - 講演依頼が来たり
  - 自信がついたり

起業したいあなたに（1/2）

Ceek.jp

- 在学中に起業するのがお勧め
  - 卒業後も継続できそう
    - 続ける
  - 失敗した！
    - 就職
- 経験の落とし穴
  - 社長の経験は社長でしか積めない（副社長と社長の溝）
  - 経験・知識にゴールはない
- IT起業であればお手軽

起業したいあなたに（2/2）

Ceek.jp

- 何かサービスを作りましょう
  - 紹介するものが無ければ営業も難しい
  - 受動的な営業を目指す
- 自身の能力を知りましょう
  - 予定時間と作業時間の記録
- 安売りは止めましょう
  - 一度下げた受託金額を上げるのは困難
- 売り切りにならないように
  - 保守契約、再利用

おわりに

おわりに

Ceek.jp

- 学生の特権
  - 失敗した際のリカバリが容易
  - 自身のブランドを持ってみませんか？
- 便利な世の中
  - IT起業であれば準備いらず
  - ウェブサービスで多人数にリーチ
- 一歩踏み出した先に
  - ワクワクする未来があるはず

とりあえず一歩前に！

「吉田光男」で検索
ceekz@mibel.cs.tsukuba.ac.jp

何かありましたらお気軽に。
@ceekz