

社会システム情報学特論

石川佳治

目次

- 情報技術の発展
- 情報爆発の時代
- 最近の話題

情報科学とは(1)

- 日本で: 情報科学 (Information Science), または, 情報工学 (Information Engineering) という呼び名が一般的
- 欧米: **Computer Science** (計算機科学, コンピュータサイエンス) という呼び名が主流
- 対象: 多岐に渡る
 - 情報と計算の理論的基礎
 - コンピュータシステムの実現手法
 - コンピュータシステムの応用



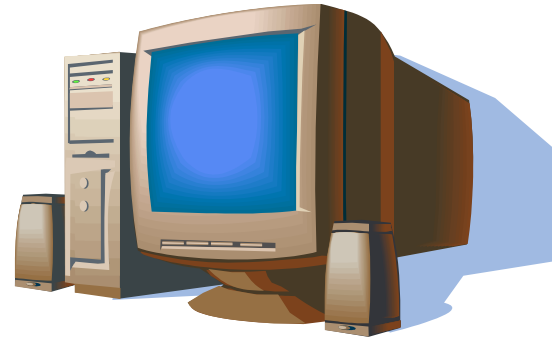
情報科学とは(2)

- 情報科学は比較的新しい学問分野
 - 1946年:最初のコンピュータENIAC登場
 - 1960年代より学問分野として確立
 - 特に技術革新のスピードが速い分野の一つ
- 学問分野の拡大
 - 当初は, コンピュータシステムをどう作るかが最重要課題
 - 今日から振り返ると限られた対象
 - 今日の情報科学:多様な拡がり
 - コンピュータを用いたさまざまな情報処理
 - コンピュータ社会を支える基盤技術



20年前を振り返ると...

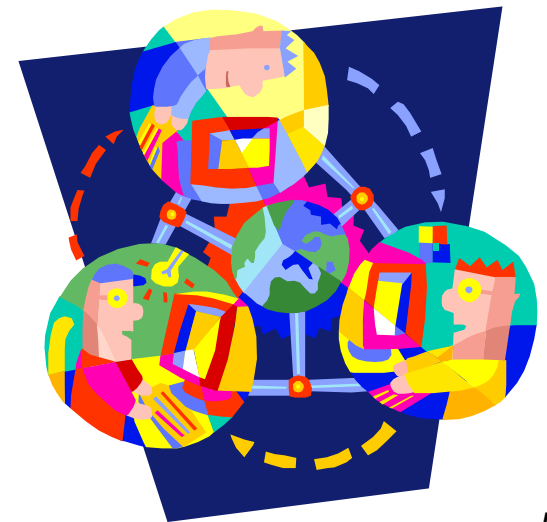
- **ドッグイヤー** (dog year)
で考えると140年前
- 1988年: パソコンが普及
しだした頃
 - PC-98シリーズが主流
- 例: PC-98 XLのスペック
 - 1986年10月発売
 - CPUクロック数: 10MHz
 - 主記憶: 1MB(メガバイト)
 - ハードディスク: 20MB
 - オペレーティングシステム: MS-DOS



- 比較: 現在のパソコン
 - CPUクロック数: 2GHz
(200倍)
 - 主記憶: 2GB(2000倍)
 - ハードディスク: 500GB
(25,000倍)

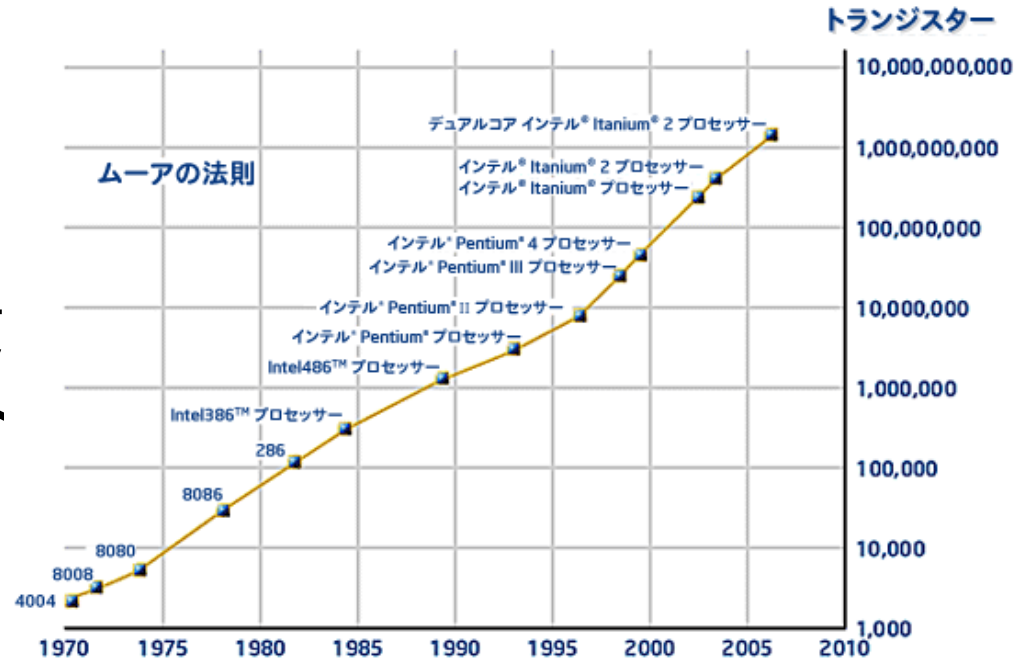
その後の歴史：ネットワーク社会への20年

- 1986年：米国で学術研究用のネットワーク基盤NSFNetが作られる
- 1980年代後半：日本の大学等でメールの利用が開始
- 1991年：Tim Berners-LeeによるWorld Wide Webプロジェクトが開始
- 1994年：Java言語発表
- 1995年：NSFNetが民間に移管⇒インターネット拡大
Microsoft Windows 95発売
- 1998年：Google創業
- 1999年：i-modeサービス開始
- 2003年：音楽配信
- 2004年：ブログ，SNSが普及

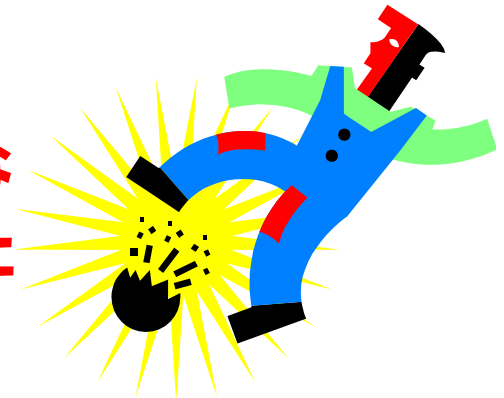


ムーアの法則(1)

- インテル設立者の Gordon Mooreにより唱えられた経験則
- 集積回路における集積度(単位面積あたりのトランジスタ数)は18ヶ月~2年ごとに**倍増**
 - 指数的な増加を意味
- 理論的裏づけはないが、これまでのコンピュータの性能向上に適合
- 技術目標としての役割

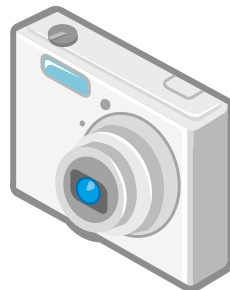
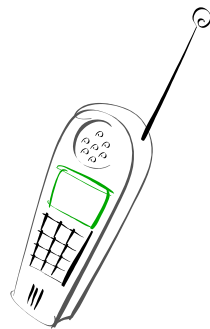


爆発的な
性能向上



ムーアの法則(2)

- 以下も大まかには(広義の)ムーアの法則に従う
 - コンピュータの速度: 単位時間に実行できる命令数
 - 価格あたりの主記憶・ハードディスクのサイズ
- 情報科学の技術も飛躍的に進展
 - 情報機器のコンパクト化・普及につながる



ムーアの法則(3)

- 今後もずっと成り立つのか?
 - ▶ 答えは**NO**: 単体のCPUには物理的な制約が存在
 - 集積回路の集積度を際限なく上げることはできない
 - 回路内を流れる電流(電子)は光の速度を超えられない
- しかし, 以下のようなブレークスルーが期待
 - 高レベルの並列処理による性能向上
 - 高速ネットワークによる多数のコンピュータの連携
 - 新しい計算パラダイムの開発
 - DNAコンピューティング
 - 量子コンピューティング

目次

- 情報技術の発展
- 情報爆発の時代
- 最近の話題

情報分野を代表する企業は？

- 20年前 : IBM



- Big Blue

- 大型計算機からPCまで

- コンピュータ(ハードウェア)を作り提供

- 10年前 : Microsoft



- Windowsオペレーティングシステム

- ソフトウェアを作りサービスを提供

- 現在 : Google

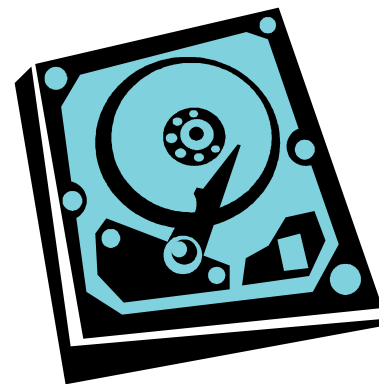


- サーチエンジン

- ウェブから情報(データ)を集めて提供

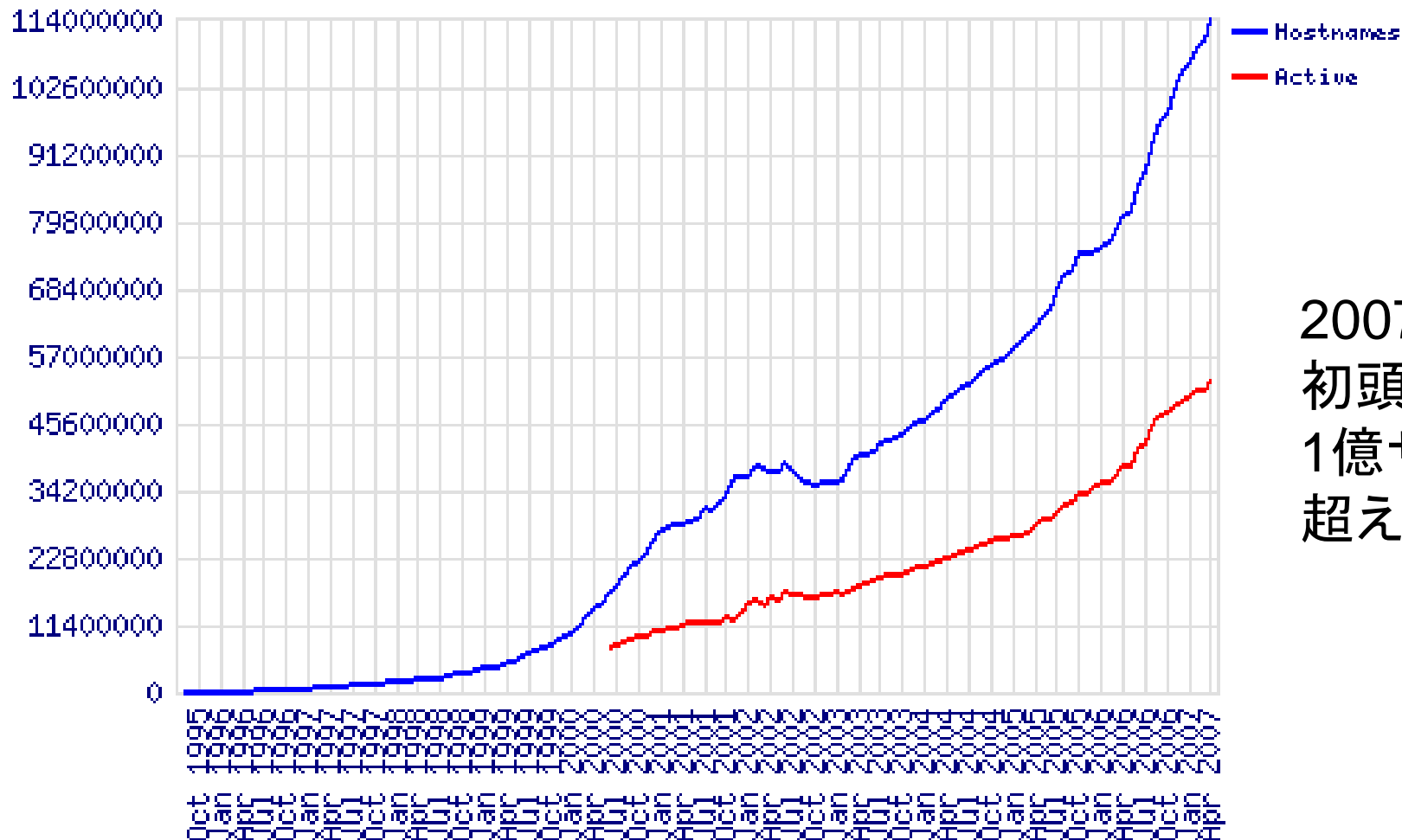
個人データの爆発

- 今日では、情報(データ)を蓄積するコストは**非常に小さい**: 今後さらに小さくなる
 - 1テラバイト(TB)のハードディスクが10万円以下
 - 新聞(文字のみ)1部は500キロバイト程度: 1TBのディスクに約5,500年分
 - iPod(80GB)には約2万曲なので、1TBには25万曲: 1曲あたり単価0.4円
- 無駄な情報を蓄積するコストよりも、整理をするコストの方が大
 - Googleデスクトップ検索の出現理由



Webの情報爆発(1)

- Webサイト数の変化(1995.10~2007.4)



2007年の
初頭で
1億サイトを
超える

Webの情報爆発(2)

- Webの総ページ数は?
 - 正確な値を求めるのは不可能: サーチエンジンではカバーしきれず, 日々新たなページが出現
 - 2005年1月時点の推計: 115億ページ
 - サンプリングと統計的な推定で求める
- Web上の新たな情報源
 - 動画配信: YouTubeなど
 - Hidden Web
 - ウェブページとして直接アクセスできないが, 検索インタフェースを通じて表示されるページ
 - 例: Amazon
 - 背後に膨大なデータベースが存在
 - ブログ, ソーシャルネットワーク(SNS)

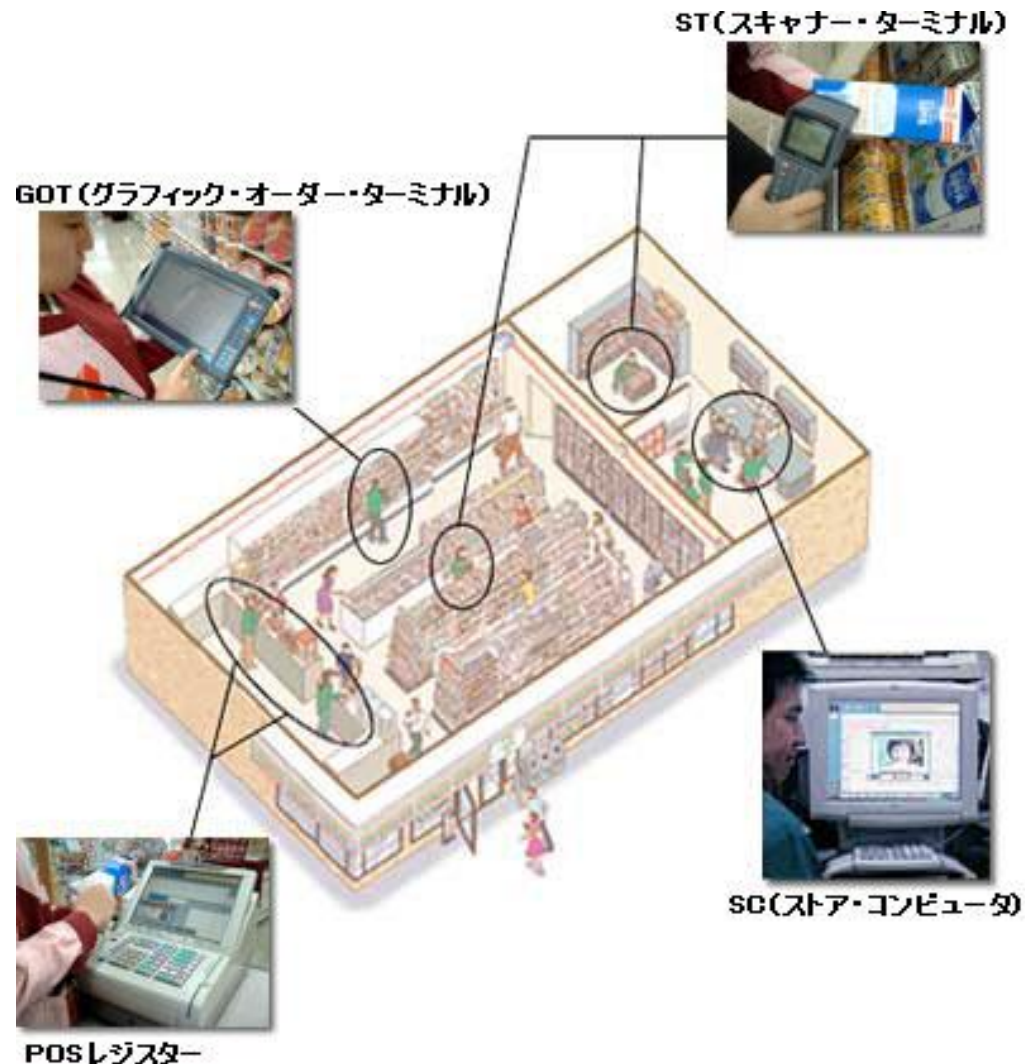


デジタルコンテンツの情報爆発

- 大量のコンテンツがネットワークを介してアクセス可能に
 - テキスト, 画像, 動画, ...
- 例: Googleブック検索
 - <http://books.google.com/>
 - 書籍の一部を画像として提供
 - 著作権が消滅している場合は, 全文を表示
- 今後さらに過去のコンテンツが閲覧可能に
 - 著作権切れの映画など

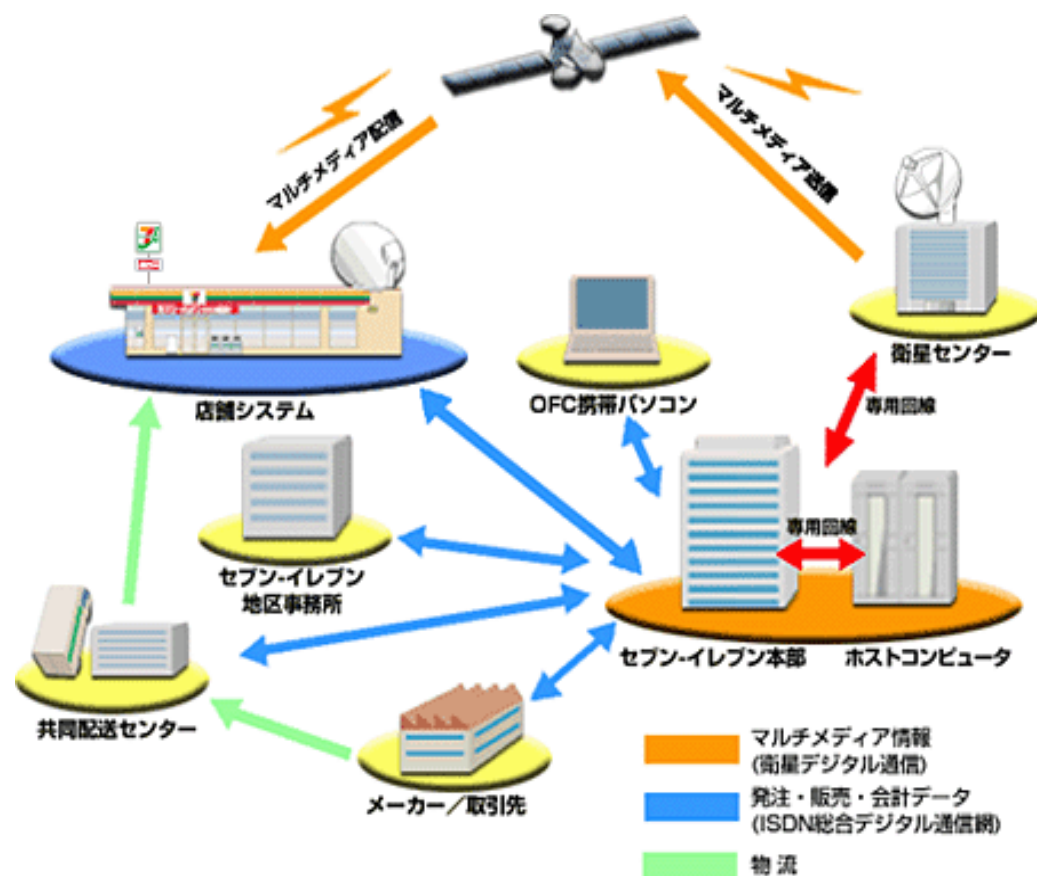
コンビニの情報システム(1)

- POS(Point Of Sales)システムにより, 購入情報をリアルタイムに収集
 - 購入品目
 - 購入日時
 - 購入者の性別, 年齢層
- 発注の精度を高め, 無駄をなくす:
「予測」に基づく発注



コンビニの情報システム(2)

- 例:セブンイレブン
- 約7,500店舗
- 各店あたり, 1日あたり約4,000件のPOSデータ:3~4GBに相当
- 単純計算で全店舗1日あたり20から30TBのデータ
- ネットワークを介して収集



センサデータ／ICタグのデータ

- センサ機器の発展
 - 無線ネットワーク技術により, センサ同士が通信
 - リアルタイムに大量の情報をモニタリング
 - 防災・防犯, 医療, 農業など多彩な応用
- ICタグの利用の拡大
 - 小型の情報チップで情報を管理
 - 物流管理, 医療, 図書館などに応用
 - トレーサビリティを実現
- ▶ **大量のデータの生成**

情報爆発時代の課題

- Question

情報爆発に打ち克つには？

- 日々大量に生み出される情報をいかに有効活用するか？
- 爆発的に増加する情報に対し、情報技術をフルに活用して打ち克つ
- 情報科学の新たなチャレンジ



目次

- 情報技術の発展
- 情報爆発の時代
- 最近の話題
 - Data Intensive (Super) Computing
 - Cloud Computing

Googleはどう実現されているのか?(1)

- 大量のデータ
 - 2007年で100エクサバイト(1000億GB)
 - クローリングしたウェブページ
 - 索引データ
 - 画像データ
- 高速な応答時間
 - 0.1秒
- 広告による収入
 - 1クリック5セント
- 詳細は以下の書籍
 - 西田圭介, Googleを支える技術: 巨大システムの内側の世界, 技術評論社, 2008年



Googleはどう実現されているのか?(2)

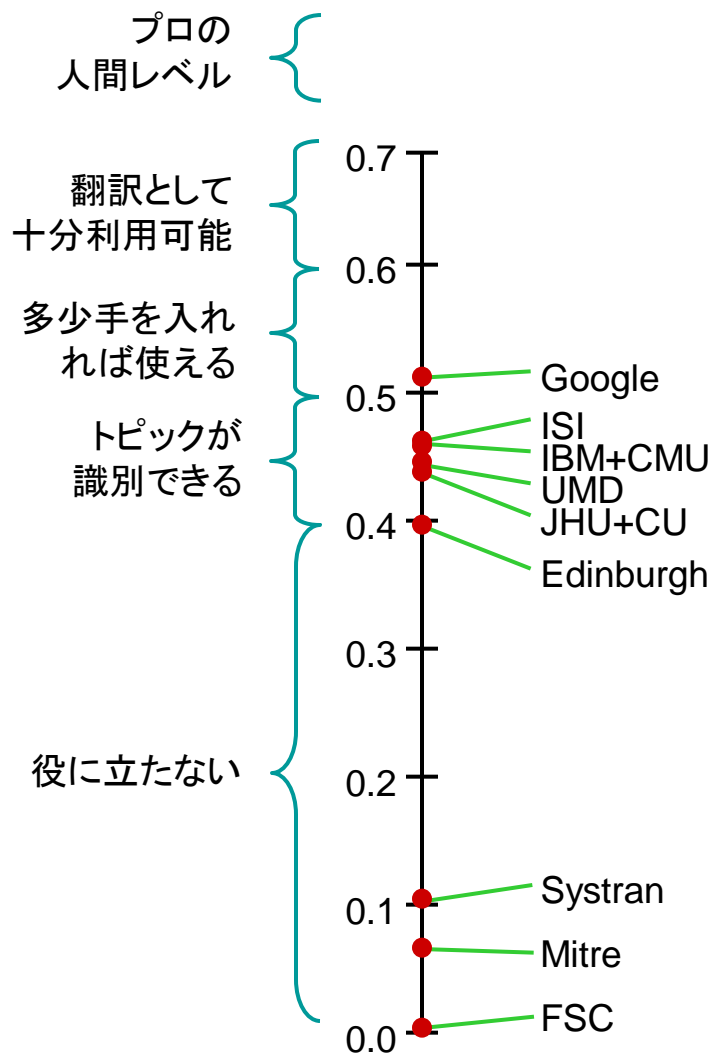
- 世界各地にデータセンターが存在
 - 分散して(一部重複)データを所有
- クラスタマシンとして実現
 - 一個一個は一般的なPC
 - 全体で300万個のCPU
- 膨大なデータの蓄積・処理技術
 - Google File System (GFS)
 - Bigtable, MapReduce, ...
 - 電力消費量の削減
- 情報科学の最先端の技術を導入

Data Intensive Super Computing(1)

- 大量のデータの蓄積・処理技術と超高性能のコンピュータで何ができるか?
 - 世界はデータであふれている
- 例:NISTによる機械翻訳のコンペティション(2005年)
 - 課題:100個のアラビア語の文を英語に自動翻訳
 - Googleは初参加
 - オーソドックスなアプローチ, ただし
 - 高レベルの翻訳例から2億語を学習
 - 単一言語からなるテキストから100兆語を学習
 - 1000プロセッサからなるクラスタマシンを利用

Data Intensive Super Computing (2)

• コンペティションの結果



より洗練されたアプローチに対して勝利

Data Intensive Super Computingの課題

- 1TBのデータを5分で処理
 - データを数百のディスクに分散
 - 数百のプロセッサで処理
 - 高速ネットワークで接続
 - 数百～万のユーザに対応
- Hadoopプロジェクト
 - Yahooによるオープンソースプロジェクト
 - Googleのような大規模データ処理機能を提供
 - <http://hadoop.apache.org/>

目次

- 情報科学とは
- 情報技術の発展
- 情報爆発の時代
- 最近の話題
 - Data Intensive (Super) Computing
 - Cloud Computing

Cloud Computingとは(1)

- インターネットを利用する人は13億人
- 何を求めている
 - アクセス:さまざまなデバイスからアクセスしたい
 - 共有:いろいろな人との情報の共有
 - 自由
 - 簡潔性, 分かりやすさ
 - セキュリティ, プライバシー

Cloud Computingとは(2)

- データは「雲」の中に
 - インターネットのどこかにあればよい
 - 例: Picasa, Gmail, Google Docs
- データへのアクセス
 - ウェブブラウザなどで可能
 - 携帯などからも利用可能
- ユーザ間での情報の共有も可能

例: Gmail



月曜: 大阪出張



水曜: 自宅から



火曜: 携帯から

by Google BETA

Welcome to Gmail

Sign in to Gmail with your Google Account

Username:

Password:

Remember me on this computer.

Sign in

[I cannot access my account](#)

@gmail.com | Settings | Help | Sign out

1 - 14 of 14

5:40 am

Weekly from The Motle... May 20

Visit http://www.travel... May 18

May 17

he Motley Fool: To prev... May 13

ctual info. Let's do ano... Apr 4

e sure we're still on for ... Apr 4

Visit http://www.travel... Mar 30

t tell me when and wh... 10/18/04

o see the pictures, so ... 9/10/04

il - Aunt Judy has acce... 9/10/04

nd Caitlin just get back... 9/9/04

ing but my scanner do... 9/9/04

w. - First off, welcome... 7/26/04

1 - 14 of 14

例: Google Docs

スプレッドシート(Excelのようなもの)を他人と共有

The screenshot shows a Google Docs spreadsheet titled "Project Orion" with a "Collaborate" sidebar open on the right. The spreadsheet contains a dashboard for "Project Orion" with various sections and data tables.

Project Orion Dashboard

Description: Project Orion is a new tool for dramatically increasing the output of widgets on an assembly line.
Release Date: 2/28/2007
Project Lead: Mary

Key Functionality Summary

Feature	Status	Priority	Owner	Comments
Robotic arm for widget creation	Green	High	2 John	In testing now
Photovoltaic sensor for quality control	Green	Medium	1 Larry	
Adapter to existing input line	Yellow	Low	0 Peter	2 weeks ahead of sched.
Assembly line apertures	Green	Low	0 Joe	
Process management console	Green	Low	2 John	

Functional Area Status

Area	Status	Lead
Product development	Yellow	John
Quality control and testing	Green	Jack
Manufacturing	Red	Boris
Distribution	Red	Sara
Sales	Yellow	Jason

Partner Status

Partner	Status	Owner
Grosman	Yellow	Sara
Hades	Green	Jason
Yestivca	Green	Sara

The "Collaborate" sidebar on the right is titled "Invite people to collaborate" and includes options to invite people as collaborators or viewers, a list of current collaborators (chris@acmesystemsinc.com, jason@acmesystemsinc.com, mary@acmesystemsinc.com, matg@google.com), and buttons to "Invite these people", "Share with everyone at acmesystemsinc.com", and "Allow anyone to view/edit".

Invite others to collaborate on the spreadsheet

Googleの次にくるものは...

