

2014年4月25日の調査報告

- 村川 – 研究室活動
- 藤井 – 全文検索
- 山地 – Subversion
- 野田 – Linked Open Data
- 碓石 – HTML5
- 田中 – 形態素解析
- 原田 – 東日本大震災

研究室活動

村川猛彦

2014年4月25日

本日本話すること

- 研究室で何をする？
- 充実した研究室活動を送るために
- 村川研究室概況

研究室で何をする？

- 学生は...
 - 修士論文・卒業論文を完成させる
 - 発表も
 - 成果を学会で発表
 - 口頭発表(予稿も)
 - 研究室内で学習・実装・評価・取りまとめ
 - 教員や他の学生と雑談(?)

研究室で何をする？

- 教員は...
 - 成果を学会で発表
 - 口頭発表, 国際会議発表, 論文誌掲載
 - 学生を指導
 - 成果やニーズをもとに, 中長期的な戦略を立て実行
 - 学内外の教員・研究者との連携も
 - 機器・書籍を購入しながら

注)短期:単年度・1年以内, 中長期:複数年度

充実した研究室活動を送るために

- 卒研究生のスケジュール
- 大学院生のスケジュール
- 研究成果と情報源の分類
- 活動心得
- 就職対策

卒研究生のスケジュール

- (3年後期: 3年配属, 情報通信システム演習)
- 3年後期~5月ごろ: 就職活動
- (3年の春休み: 卒研着手判定)
- 随時(教員の指示で): 進捗報告
- 8月下旬: 院試(筆記, 面接)
- 2月中下旬: 卒論提出・発表

大学院生のスケジュール

- 半期に2回以上：大ゼミ発表
 - 大ゼミはスライド準備だけでなくレジュメ作成も
 - M2の後期2回目は修論中間発表
- 随時(教員の指示で)：進捗報告
- 年度に1回程度：学会発表
- M1後期～M2の5月ごろ：就職活動
- M2の2月上中旬：修論提出・発表

注) 大ゼミ発表回数は知的モデリングクラスタの場合
(クラスタにより異なる)

研究成果と情報源の種類

- 学位のための論文
 - 博士論文, 修士論文, 卒業論文
- 学会の刊行物
 - 学術論文, 国際会議予稿集, 研究会・シンポジウム, 全国大会
- その他(参考文献に入れてよいもの)
 - 紀要, 報告書, 白書

参考文献はなぜ必要？

→先人の積み重ねた発見に基づいて, 新たな発見を得るため

活動心得

- 新規性は先行事例(文献)を踏まえてこそ
- 教員は門番ではなく仲間
- データこそ宝, アイデアは包み隠すな
 - 普段から記録を付けること
- 発表準備は「2週間前から」「2日前に発練」

就職対策

- M2と4年は就職活動を優先してよい
その分、陰で研究活動を怠らずに
 - 4年は、就職と進学を秤にかけてほしい
- アドバイス
 - 情報の鮮度と正しさを常に確認
 - しっかり準備(学内でも)
 - 笑顔で、礼儀正しく、はっきり話す

村川研究室 概況

- 略歴

- 1999年4月：中川研究室に参加
- 2006年10月：村川研究室発足，3年生受入

- 特色

- 理念：尊重・修得・蓄積
- 主な研究テーマ：①情報検索・推薦②データベースシステム

- 強み：ソフトウェア活用（特にLinux・検索エンジン）

今年度の研究の方向性

- 利用するデータ(コンテンツ)に注意し、利用者を支援するシステムを開発したい！
 - 情報推薦システム
 - 災害記事データベース
 - 見守りサポーター支援 など
- 基礎となるのは、テキスト処理
 - システムに応じて、「データベース」「全文検索」も
 - 「PHP禁止」に協力ください

関連情報

- 村川研究室とは
 - <http://www.wakayama-u.ac.jp/~takehiko/lab.html>
- 俺流解説・大学編
 - <http://d.hatena.ne.jp/takehikom/20140421/1398083013>
- 巨人の肩の上 - ウィキペディア日本語版

調査報告 全文検索について

2014/4/25

村川研 B4

藤井 浩平

全文検索とは

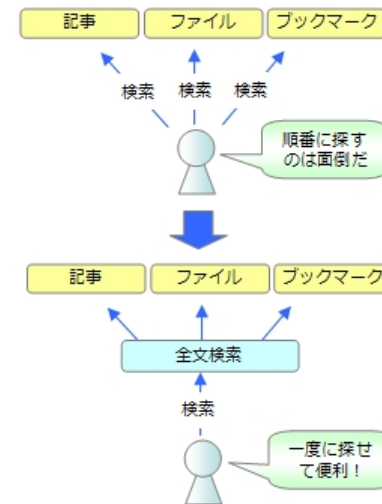
- ▶ 検索対象となるデータ全体に対して検索を実行すること
 - ▶ 例えば、フォルダに複数の文書ファイルが入っている場合、「コンピュータ」というキーワードで全文検索すると、「コンピュータ」という言葉が入っているすべての文書のすべての箇所をピックアップしてくれる

➤ メリット

- ▶ 検索範囲としては最も広い
- ▶ 最も「漏れ」が少ない

➤ デメリット

- ▶ 細かい条件を指定しにくく、絞込みにくい



全文検索の主な用途

- WWW検索サービス
 - ▶ 検索サービスの中では、超大型の機能が求められる分野
 - ▶ 現在では「Google」または「Bing」のいずれかに集約されつつある
- 企業向け社内検索サービス
 - ▶ 社内ファイルサーバの文書資産を高速全文検索するシステム
 - ▶ WordやExcelなどのオフィスから、メール、データベースなどの多くのファイル形式に対応する
- デスクトップ検索
 - ▶ 個人のローカルファイルを検索するためのアプリケーションソフトウェア
 - ▶ Word、Excel、PDFなど様々なファイル形式に対応している

代表的な全文検索エンジン (1)

- Hyper Estrair
 - ▶ わかち書き方式も併用可
 - ▶ 標準添付のプログラムが充実
 - ▶ 大規模なインデックスも作成可
 - ▶ mixiの検索に使われている

- ▶ Namazu
 - ▶ わかち書きベース
 - ▶ 古くからあり、日本で広く使われている全文検索システム
 - ▶ 小中規模を対象とする

代表的な全文検索エンジン (2)

- ▶ Lucene
 - ▶ N-gramやわかち書き形式でのインデックス作成ができる
 - ▶ Javaによる全文検索システム
 - ▶ Wikipediaの検索に使われている
- ▶ Senna
 - ▶ わかち書き・N-gramベースの選択
 - ▶ MySQLの中に全文検索エンジンを組み込むパッチが提供されている
 - ▶ 2ちゃんねるの検索に使われている

検索エンジンの比較（1）

▶ ディレクトリ型検索エンジン

- ▶ このタイプの検索エンジンは、集められたWebページを一定のルールで評価し、そのルールにあったページを、ユーザの使いやすいような領域に分類して、登録する
ユーザは、その領域（カテゴリー）を選びながら目的のページを探し出す

- ▶ このような検索方法を、カテゴリー検索という

- ▶ ディレクトリ型検索エンジンとは、日本ではYahoo! Japan に代表される

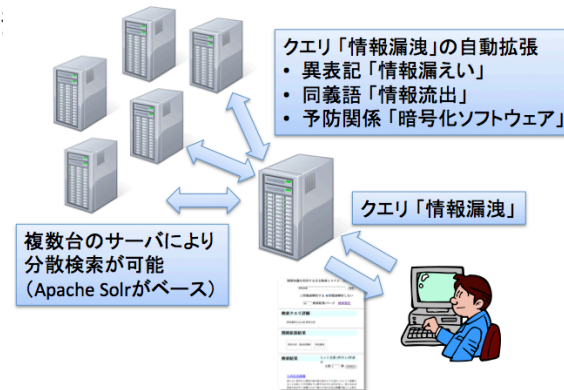


検索エンジンの比較（2）

- ▶ 全文検索型検索エンジン（ロボット型検索エンジン）
 - ▶ このタイプの検索エンジンは、集められた膨大な量のWebページに含まれる言葉をほぼそのままインデックスし、キャッシュする
 - ▶ ユーザは、探しだしたい情報に関連する言葉を使い、検索する
 - ▶ このような検索方法を、キーワード検索という
 - ▶ 代表例としては、Googleが挙げられる

関係知識を利用する全文検索システム

- ▶ QE4Solrとは
 - ▶ 異表記辞書や同義語辞書、単語間意味的關係知識等を用いて、クエリを自動拡張する全文検索システム
 - ▶ 大量の異表記、同義語、意味的關係知識によりクエリを拡張することで、検索の漏れの防止や、意外だが有用な情報の発見が期待できる
 - ▶ 索引付けと検索の並列処理が可能で、Webアーカイブ等の大規模な文書データも効率的に処理で



具体例

▶ 異表記, 同義語の例

- ▶ (ギョウザ, ギョーザ)、(バイオリン, ヴァイオリン)、
(情報漏洩, 情報漏えい)、(ろうきん, 労働金庫)、
(白滝, 糸こんにゃく)、(キャリアサポート, 就職支援)

▶ 意味的關係知識の例

- ▶ (原因－結果) 關係の例：(動脈硬化－脳梗塞)、
(トラウマ－PTSD)、(断層－直下型地震)
- ▶ (トラブル－予防策) 關係の例：
(情報漏洩－暗号化ソフトウェア)、(床ずれ－エアマット)

全文検索以外の検索方法

- ▶ SQLでのテーブルレコードの抽出
 - ▶ SQLはクエリの本になっているもので、クエリで基本となるものは、選択クエリ
 - ▶ 選択クエリとは、テーブルレコードを抽出したり、検索を行ったりするクエリの本になるもの
- ▶ 基本構文
 - ▶ 選択クエリの本形は、SQLの本形と同じ
 - ▶ SELECT フィールド名 FROM テーブル名;

具体例

▶ サンプルテーブル（学生情報）

学籍番号	氏名	生年月日	出身地	学年
60161111	和歌山 一郎	1992/1/1	和歌山	4
60162222	大阪 二郎	1992/2/2	大阪	4
60163333	京都 三郎	1992/3/3	京都	3

▶ サンプル構文

- ▶ `SELECT 学籍番号, 出身地, 学年 FROM 学生情報;`
- ▶ 全てのフィールドを対象にするなら、各フィールド名を略しアスタリスク（*）を使用することができる

レコードの抽出

- ▶ テーブルレコードをある特定の条件で抽出することができ、この場合、WHERE句で抽出条件を指定する

- ▶ SELECT テーブル名 FROM フィールド名 WHERE句 条件句

- ▶ 条件句に用いる比較演算子

記号	意味
=	等しい
<>	等しくない
> =	より大きいか等しい
>	より大きい
< =	より小さいか等しい
<	より小さい
AND	かつ
OR	または

具体例

▶ サンプルテーブル（学生情報）

学籍番号	氏名	生年月日	出身地	学年
60161111	和歌山 一郎	1992/1/1	和歌山	4
60162222	大阪 二郎	1992/2/2	大阪	4
60163333	京都 三郎	1992/3/3	京都	3

▶ サンプル構文

▶ `SELECT * FROM 学生情報 WHERE 学年 >= 4;`

あいまいな条件抽出

- ▶ LIKE演算子を使った条件検索
 - ▶ WHERE句で検索条件を指定するときに、LIKE演算子を使用すると、あいまい検索を行うことができる
 - ▶ LIKE演算子は、ワイルドカードを使って文字列を検索するための演算子
 - ▶ LIKE演算子とワイルドカードを用いた検索は「あいまい検索」とも呼ばれ、幅広い対応が可能

あいまいな条件抽出

▶ 主なワイルドカード

ワイルドカード	説明	事例
アスタリスク (*) パーセント (%)	0文字以上の任意の文字列を表す	名前 LIKE "鈴木*" (鈴木以下の文字は何でも検索する)
疑問符 (?) アンダスコア (_)	任意の一文字を表す	名前 LIKE "鈴木?" (鈴木以下の1文字は何でも検索する)
シャープ (#)	任意の半角の1数字を表す	数字 LIKE #32 (132、332などが該当する)
[文字リスト]	文字リストに含まれる全角、または半角の1文字	名前 LIKE "[田中]" (田、中が該当する)
[文字A - 文字B]	文字A～文字Bの範囲の全角または半角の1文字	名前 LIKE "[A - C]" (A B Cが該当する)

具体例

▶ サンプルテーブル (学生情報)

学籍番号	氏名	生年月日	出身地	学年
60161111	和歌山 一郎	1992/1/1	和歌山	4
60162222	大阪 二郎	1992/2/2	大阪	4
60163333	京都 三郎	1992/3/3	京都	3

▶ サンプル構文

- ▶ `SELECT * FROM 学生情報 WHERE 氏名 LIKE "*和歌山*";`
- ▶ `SELECT * FROM 学生情報 WHERE 氏名 LIKE "*[和歌山]*";`

まとめ

- ▶ 全文検索とは、検索対象となるデータ全体に対して検索を実行することで、検索範囲が最も広く、漏れが少ない
- ▶ 一般的なGoogleの検索や、企業の社内検索に使われている
- ▶ 様々な全文検索エンジンがあり、Wikipediaや2ちゃんねるなど様々なところで活用されている

Subversion

4月25日

B4 山地康太

Subversion とは?

- Subversion は、複数の開発者が共同で一つのソフトウェアを開発する際などに、ソースコードやドキュメントを管理するのに用いられる。
- フリー・オープンソースのバージョン管理システムである。時間とともに変化するファイルやディレクトリやフォルダを管理する。
- このため、古いバージョンのデータを戻したり、変更履歴を確認したりすることができる。
- CVS(Concurrent Versions System)を参考に開発され、基本的なコマンドなどはCVSと共通になっている。

Subversion の特徴

- 1.バージョン管理システムを使い履歴を管理。
- 2.コピー・変更・マージモデルモデルを使いファイルを共有
3. バイナリファイルの履歴も効率的に管理できる。
- 4.ファイルやフォルダのリネームや移動も履歴として追跡できる。

1.バージョン管理システム

- ファイルの変更履歴を記録するソフトウェア
- 繰り返し編集を重ね変化して行くファイルの内容を記録し、後から特定のバージョンを取り出したり、いつ、誰が、何の目的で、どう修正を加えたかを調べることが可能になる。

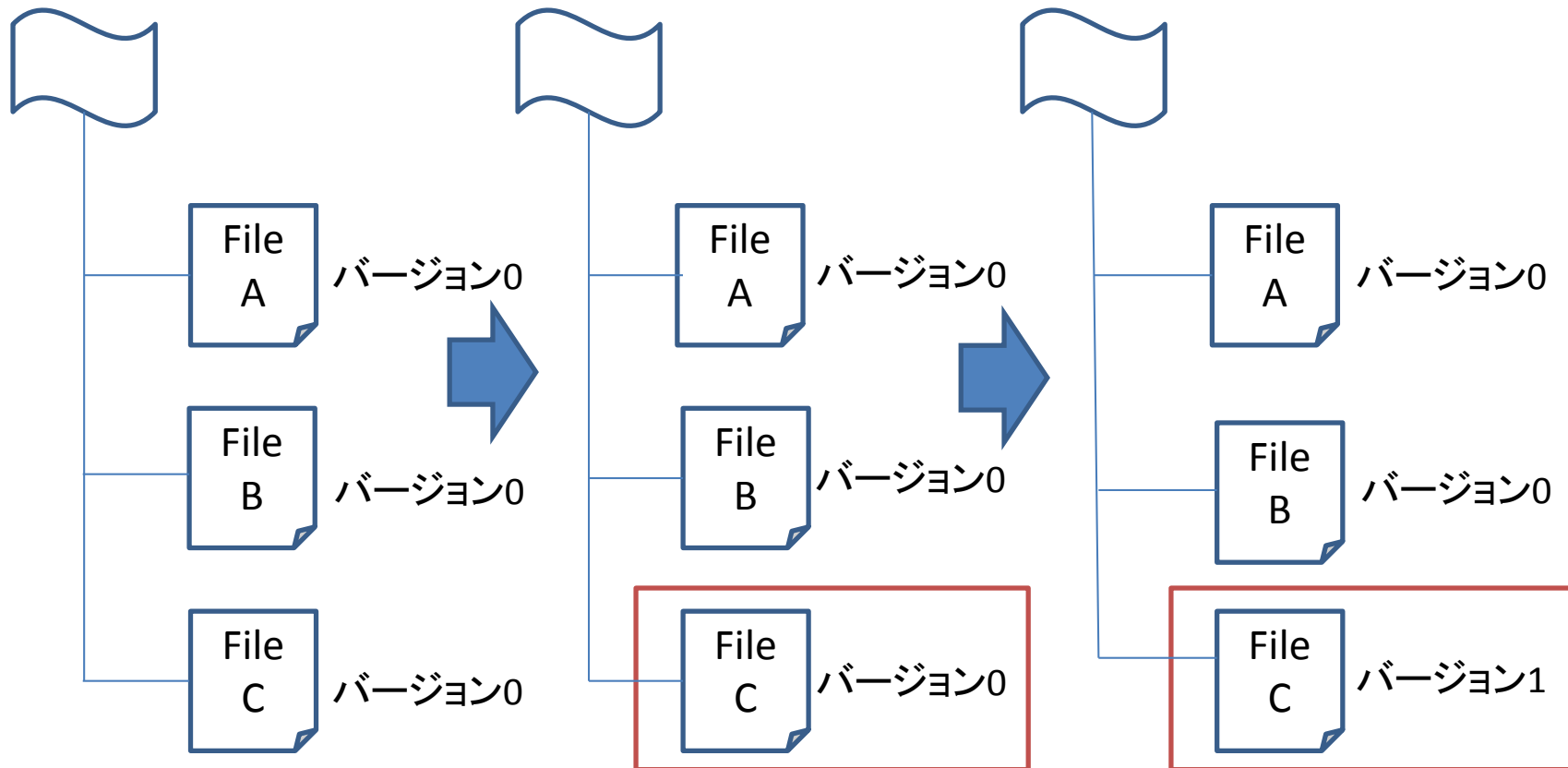
バージョン管理システムの管理方式

バージョン管理システムでは、ファイルの各バージョンをデータベースに保持しており、このデータベースを一般にリポジトリと呼ぶ。

1. ファイルをリポジトリに登録する。
2. ファイルをリポジトリからローカル環境に取り出す
(チェックアウト)
3. ローカル環境で、ファイルに対し変更を行う。
4. 変更したファイルをリポジトリに書き戻す(チェックイン)

ファイルがチェックインされると、システムによって「いつ」「誰が」「どんな変更を行った」等が記録され、後から参照できる。また必要に応じて古い版を取り出すことも出来る。

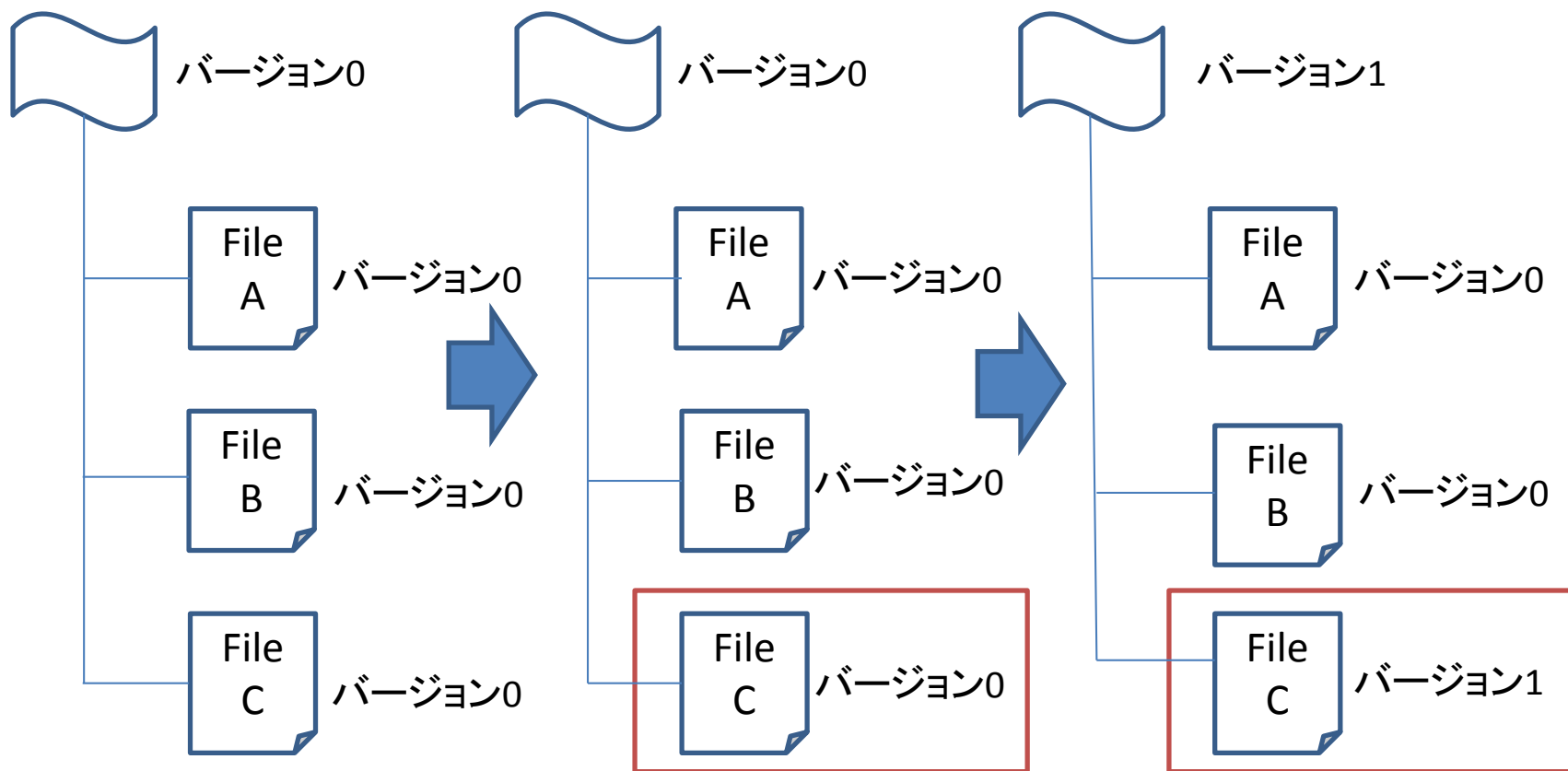
CVSでのバージョン管理の仕方



「File C」をコミット

「File C」のバージョンが1に変更される。

Subversionでのリビジョン管理の仕方



「File C」をコミット

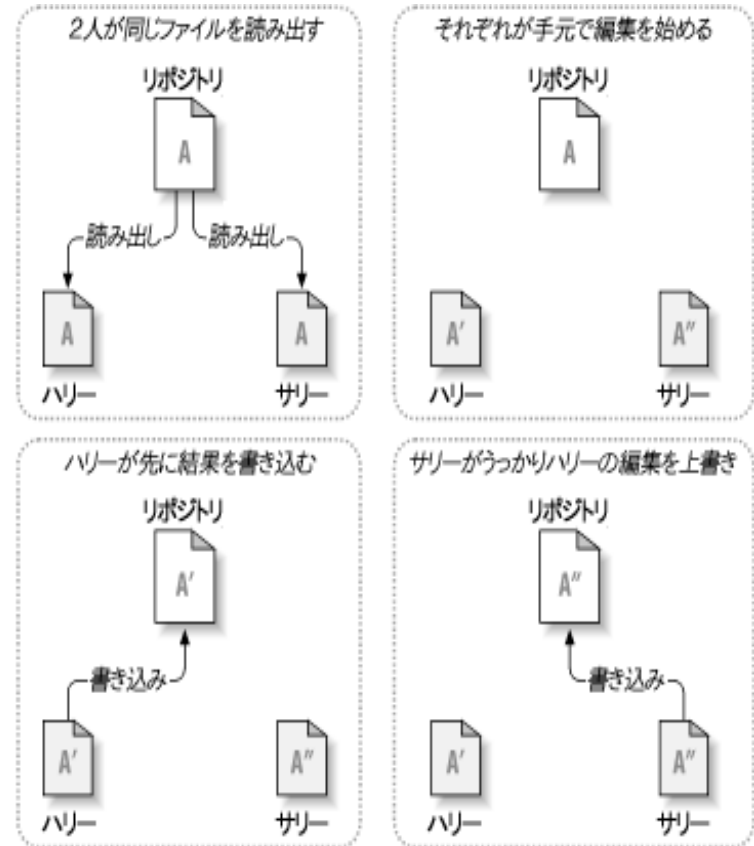
「File C」のバージョンが1に変更され、リポジトリのバージョンも1になる。

2. バイナリファイルの管理

- CVSではバイナリファイルは「完全なコピー」として格納される。
- Subversionでは「テキストファイル」「バイナリファイル」の違いに関係なく、バイナリ差分抽出アルゴリズムを使い、ファイル間の違いを差分として管理する。

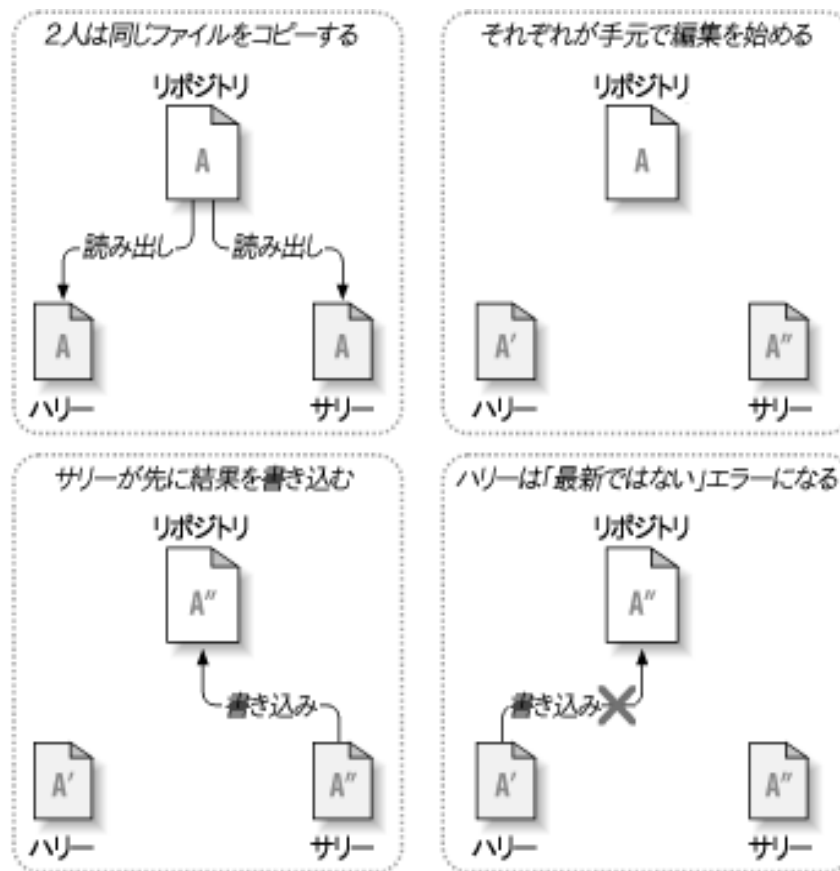
3.ファイル共有の問題

どのようにユーザーに情報の共有をさせつつ、偶然にも他人の邪魔をしないようにするかが問題です。
リポジトリ内の他人の変更を、誤って上書きしてしまうことは容易に起こりうることである。



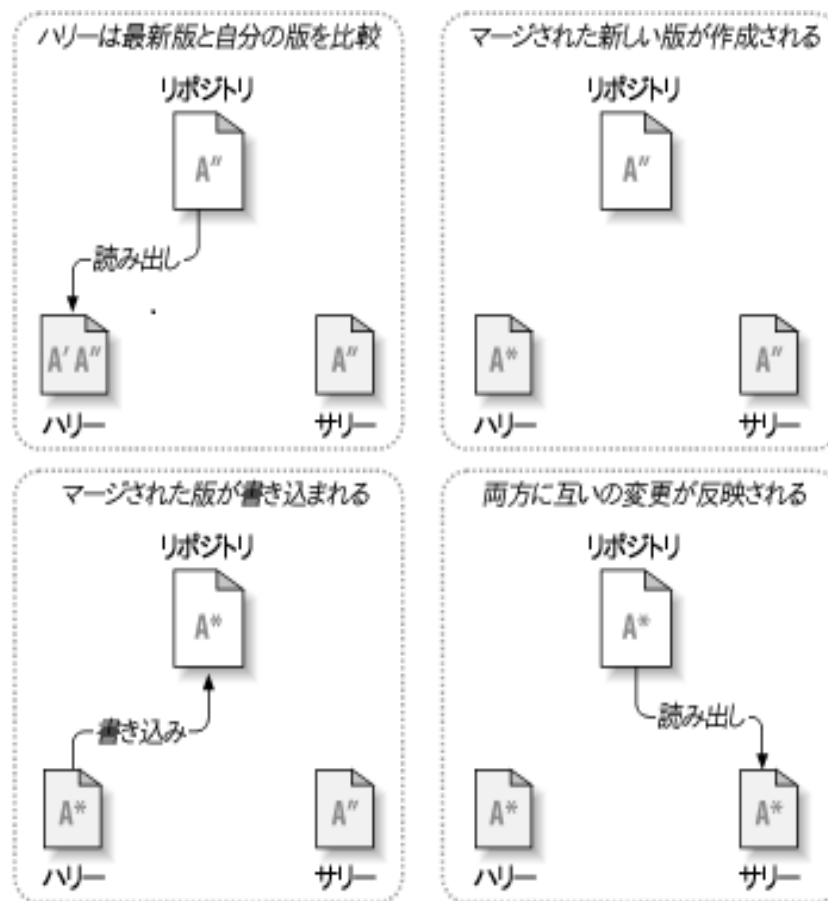
コピー・変更・マージモデル

- ユーザーのクライアントが個別にリポジトリを読み込み、ファイルやプロジェクトの個人的な作業コピーを作成する。
- ユーザーは並行して作業し、個人のコピーを変更する。
- 最後に個人のコピーを新しい最終版にマージする。



コピー・変更・マージモデル

- 最終版にマージする際に他の人も変更している場合がある。
- ファイルの変更が重ならない場合には両方の変更点を統合します。
- 変更点が重なっている場合を競合といい、競合が起こると競合した変更を両方とも見ることができる。
- ソフトウェアは自動的に競合を解決できない



メリット

- 誰がいつどのように変更したかが分かるので複数人で開発をするときにいちいち連絡を取り合う手間が省ける。
- 色々な種類の形式のファイルを管理でき、バイナリファイルでも元に戻せる。
- 同じファイルを触っていたとしても間違っても上書きすることなく、自動で書き換えたり競合を知らせてくれる。

Windows,Linuxでの使い方

1.リポジトリの作成

2.ファイル・ディレクトリ(フォルダ)をリポジトリに追加

3.チェックアウトのためのフォルダを作成

4.3で作成したファイルにリポジトリに登録したファイルをコピー(チェックアウト)し編集

5.リポジトリへ編集したファイルを登録

Linked Open Data (LOD)

村川研究室

野田 長寛

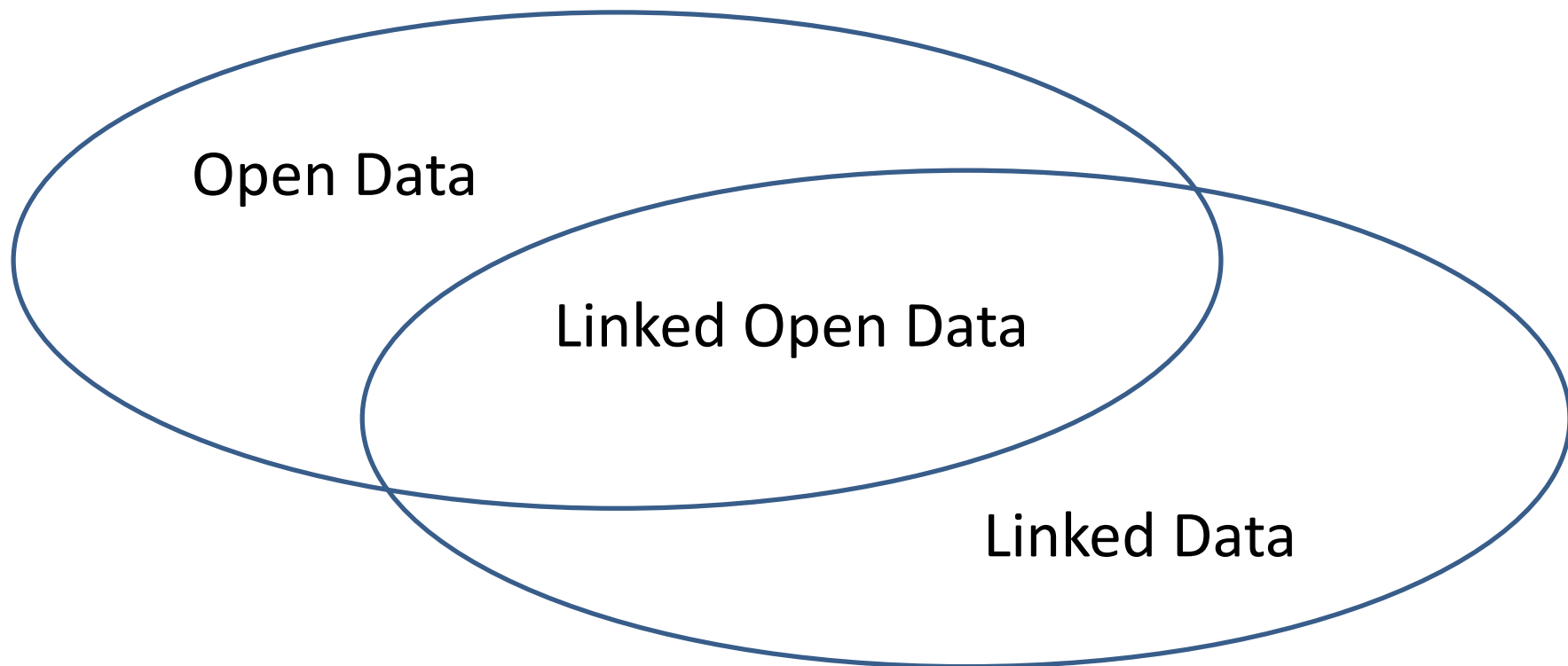
背景

- ウェブサイトとウェブサイト間のリンクを中心に広がってきている「World Wide Web」は、現代社会において欠かすことのできないものとなっている。



- WWWが持つ性質であるデータの「公開」と「共有」をより効率的かつ実践的に行えるようにするために、「Linked Open Data」が発表された。

Linked Open Data (LOD)のはじまり



Open Dataとは

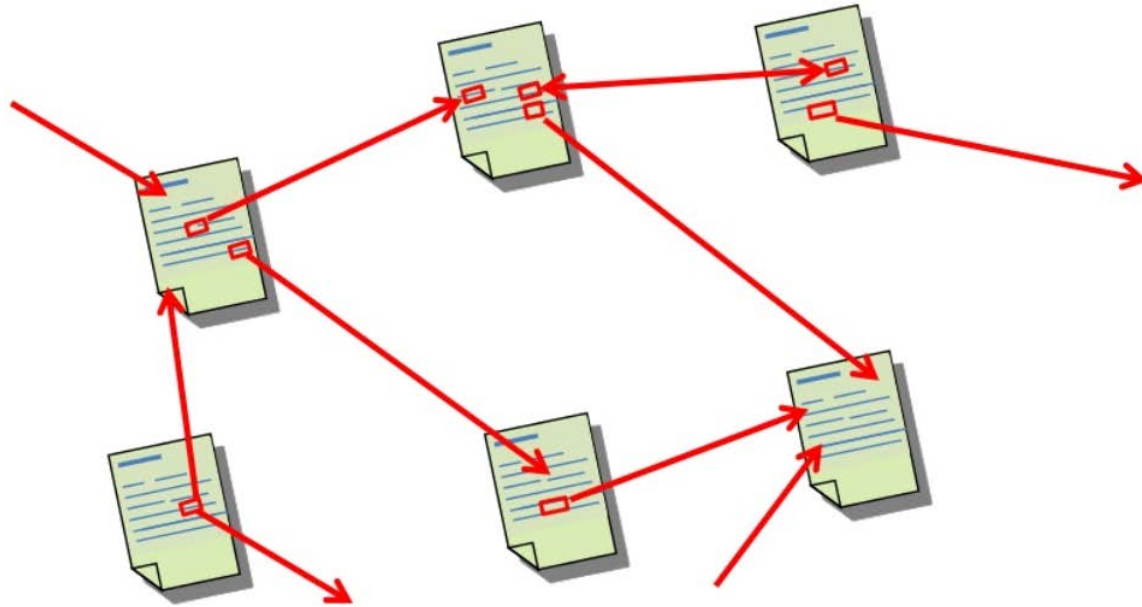
- 政府・自治体・企業などの組織または個人が所持しているデータを公開し、第三者が自由に再利用でき、かつ再配布することができるデータのことである。

<オープンライセンス>

- 情報を最小限の制約以外で自由に使うことを許すライセンス。

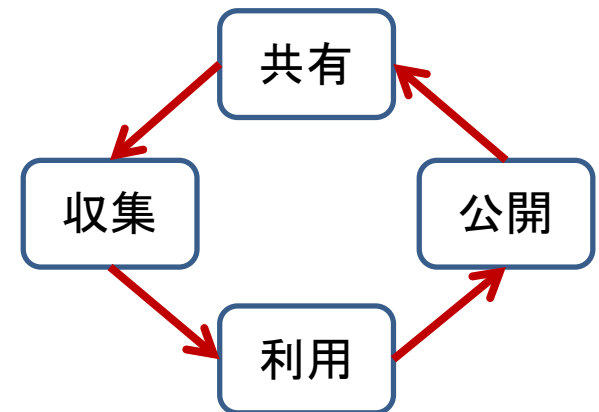
Linked Data

- World Wide Web上で、異なる情報源のデータを型付リンクによって結びつけることである。



Linked Dataの役割

- 収集→利用→公開→共有のよりよいループの完成.
- 公開
 - 人が使いやすい形にする.
 - コンピュータが使いやすい形にする.
- 共有
 - 相互にリンクし合う.
 - 概念を共有する.



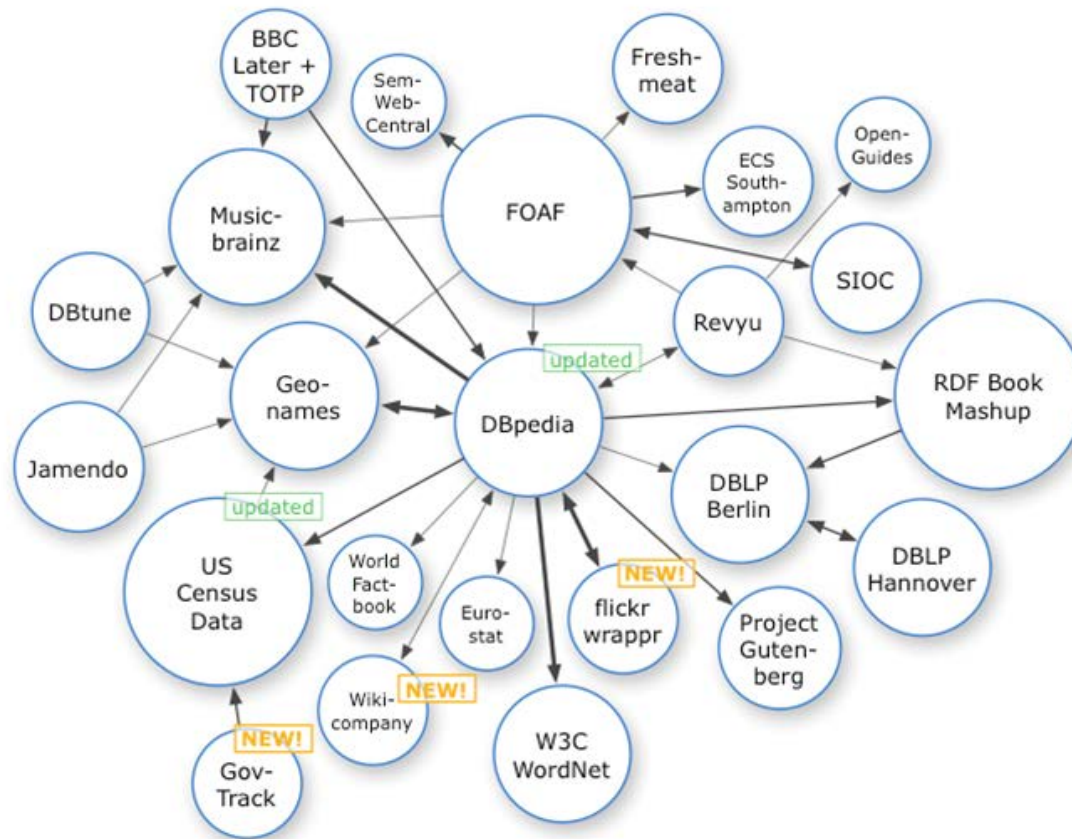
Linked Open Data (LOD)とは

- Linked Open Dataとは、ウェブ上でコンピュータ処理に適したデータを公開・共有するための技術であり、また方法論の総称でもある。

LODに関する4つの原則

1. あらゆるデータの識別子として, URIを使用する.
2. 識別子には, HTTP URIを使用し, 参照やアクセスを可能にする.
3. URIにアクセスされた際には, 有用な情報を標準的なフォーマットで提供する.
4. ウェブ上の情報発見を支援する(関連情報のリンクを含める).

LODクラウド図



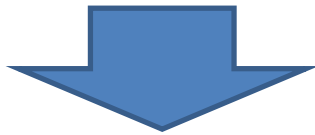
Cygri (2007.9) 「Datasets in the Linking Open Data project, as of September 2007」
<http://ja.wikipedia.org/wiki/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB:Linking-Open-Data-diagram_2007-09.png> (2014.4.24 アクセス)

LODの関連項目

- ウェブマイニング
- データジャーナリズム
- ビッグデータ
- 地理情報システム

ビッグデータ

- スマートフォンを通じて、個人が発する情報など、日々生成させるデータの集合を差し、膨大なだけでなく、リアルタイムに増加・変化する多様で複雑なデータのことである。



- 保管して、解析することで、ビジネスや社会に有用な知見を得たり、新たな仕組みや社会システムを生み出すことができる。

HTML5についての調査報告

2014年4月25日

村川研究室 M1 碓石浩文

目次

- HTMLとは
- HTMLの特徴
- 既存のHTMLとの違い
- HTMLの宣言
- HTMLのブラウザ対応状況

HTMLとは

- HyperText Markup Languageの略称
- ウェブ上の文章を記述するためのマークアップ言語
 - マークアップ言語とは文章の一部をタグで囲うことで文章の構造, 大きさなどの修飾情報を文章中に記述していく記述言語
- スイスのCERN(欧州原子核研究機構)のティム・バーナーズ・リーによって考案される

HTMLの特徴

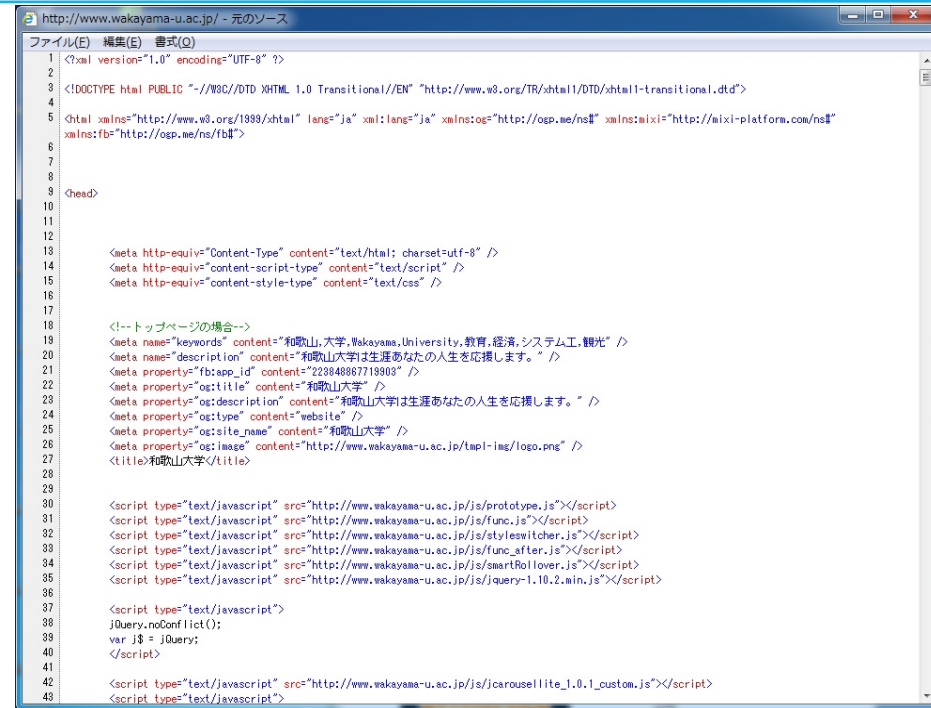
- 他の文章にリンクを作成できる
- タグによる「段落」、「強調」、「改行」などができ、ページを分割することもできる
- 画像を貼り付けることができる
- 表や図の作成ができる

HTML文章とその構造

HTMLサンプル



ソースコード



転載元URL: <http://www.wakayama-u.ac.jp/>

既存のHTMLとの違い

- 要素・属性の追加と廃止
- 文書構造の表示(文書構造を示す)
- 入力補助やチェック機能の充実
- 動画や音声データの使用
- APIの追加

要素・属性に関して

- 追加されたもの

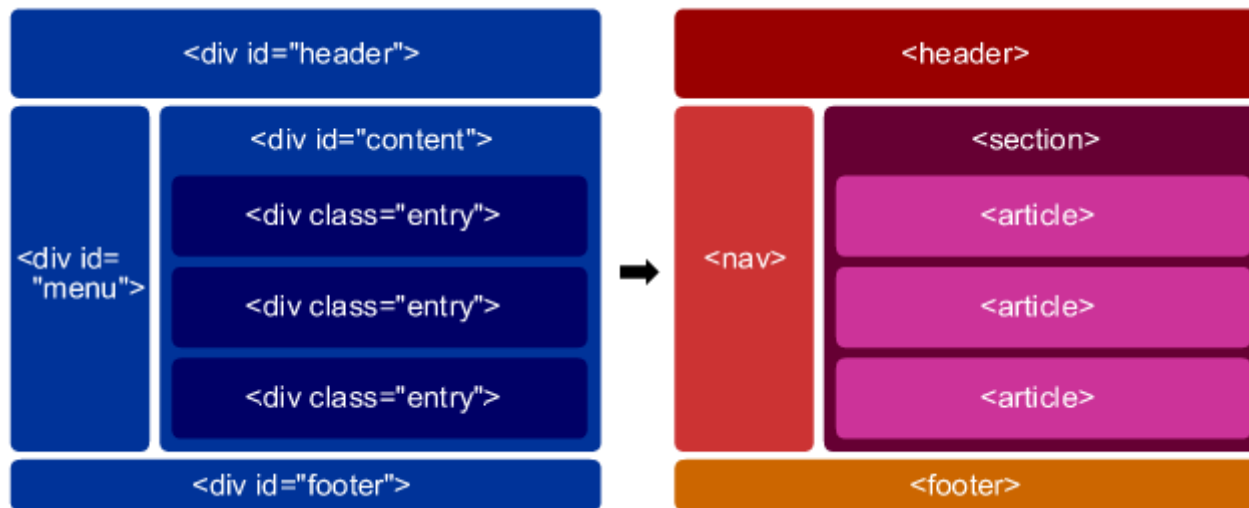
- `<audio>`……音声を再生できる
- `<header>`……ヘッダであることを示す
- `<footer>`……フッタであることを示す
- ……
- ……

- 廃止されたもの

- `<center>`……センタリング(中央ぞろえ)
- ``……フォントの種類, 大きさ, 色の指定
- `<dir>`……リストを表示する
- ……
- ……

文章構造

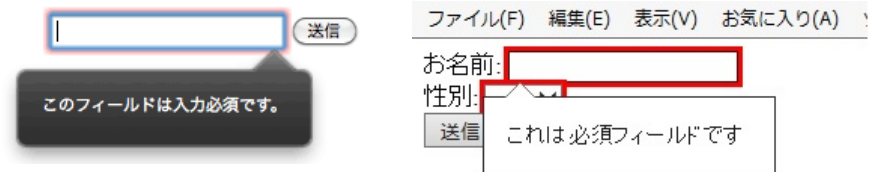
適切な役割の要素が追加されたことにより文書構造がより明確に伝えられるようになっている



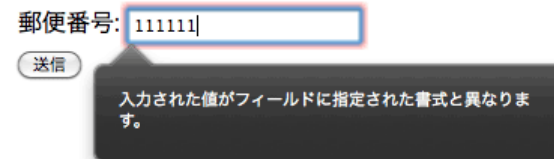
画像引用元: <http://www.htmq.com/html5/002.shtml>

入力補助、チェック機能

- テキストなどを入力したとき、エラーなどのコメントが表示される



- 数字や文字列などの入力値の指定



- 入力値の最大・最小を制限できる

画像引用元 : <http://www.atmarkit.co.jp/ait/articles/1104/25/news137.html>

動画、音声データ

- HTML5では動画や音声データを扱える.

以前のバージョンでは動画や音声をWebページに埋め込む際にはFlashなどを使用していたが、

新しく追加された<video>や<audio>といったタグが追加されたことにより、<image>で画像を使

用する感覚で動画や音声を使用できる.

APIについて

API(Application Programming Interface)
の追加により,

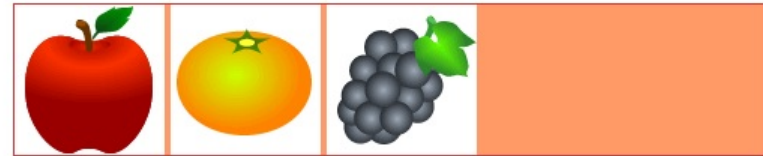
- ユーザ位置情報の取得
- ドラッグ & ドロップ機能

などなど・・・

が使用できるようになっている

画像引用元：<http://www.htmq.com/dnd/sample/001.html>

好きな果物は何ですか？



下のボックスに入れてください。



好きな果物は何ですか？



下のボックスに入れてください。



HTMLの宣言

HTML5以前の場合

- HTMLのバージョンについては<html>より以前に宣言する必要がある, Strict(厳密型),

Transition(移行型), Frameset(フレーム設定型)の3種類がある

- 以下はHTML4.01の宣言

- Strict

- <!DOCTYPE HTML PUBLIC "-//W3//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">

- Transitional

- <!DOCTYPE HTML PUBLIC "-//W3//DTD HTML 4.01Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">

- Frameset

- <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN" "http://www.w3.org/TR/html4/frameset.dtd">

HTMLの宣言

HTML5の場合

宣言

➤ <!DOCTYPE html>

- 宣言をしないと、過去のブラウザとの互換のため、

異なった解釈で表示されることもある

ソースコード例

```
<!DOCTYPE html>
< html>
< head>
< meta charset="UTF-8">
< title>HTML5サンプル</title>
< /head>
< body>
< p>HTML5で作成しました！ </p>
< /body>
< /html>
```

HTML5の対応ブラウザ状況

HTML5が登場しているが、ブラウザが対応していなければWebページの閲覧ができない

主要ブラウザのHTML5対応バージョン		リリースした時期
Internet Explorer	Internet Explorer 8以降	2009年3月
Firefox	Firefox 3.5以降	2009年6月
Chrome	Chrome 3.0以降	2009年9月
Safari	Safari 3.1以降	2008年3月

形態素解析について

2014/04/25

村川研究室

60120075

田中 勝

形態素解析とは(1)

- 計算機を用いた自然言語処理の基礎技術のひとつ
- 与えられた文を形態素単位に区切る処理
 - 形態素とは？
 - 意味を持つ最小の言語単位
 - 「辞書にある文字列」と「それに付与されている品詞など」の組
 - 形態素解析を行うには、この形態素を定義した辞書が必要
 - わかち書きをしない日本語も区切ることができる
 - わかち書き: 文や文章をわかりやすくするため、語と語、あるいは文節と文節の間をあけて書くこと

形態素解析とは(2)

- 与えられた文を形態素単位に区切る処理(続き)
 - 形態素解析の手順
 - (辞書を見ながら)文から単語を切り出す
 - (辞書にある)単語と品詞の組を当てはめていき, 文法的に正しい品詞の並びを探す
 - 文法的にありえない品詞の組み合わせは除外していく
 - 英語(空白による区切りがある)
 - This is a pen → This / is / a / pen → This(代名詞) / is(動詞) / a(冠詞) / pen(名詞)
 - 日本語(区切りがない)
 - 和歌山大学は静かな場所だ
 - 和歌/山/大学/は/静/かな/場/所/だ, ..., 和歌山大学/は/静か/な/場所/だ, ...
 - 和歌山大学(名詞)/は(助詞)/静か(形容詞)/な(助動詞)/場所(名詞)/だ(助動詞)

形態素解析の役割

- コンピュータによる言語の理解
 - 構文解析, 意味解析, 文脈解析, などの高度な自然言語処理に利用
 - 人工知能と言語学の一分野
- 自然言語処理システムとしてそのまま利用
 - 全文検索システムなどのインデックス作成
 - 日本語入力システム(かな漢字変換: ひらがな→漢字仮名交じり文)
 - MS-IME, ATOK, Google日本語入力
 - テキスト読み上げソフト(漢字仮名交じり文→ひらがな)
 - AquesTalk
 - 言語統計のための前処理

形態素解析エンジン

- 入力文を単語単位に分割し品詞を付与するツール
- 日本語の形態素解析エンジン
 - MeCab(和布蕪)
 - ChaSen(茶釜)
 - JUMAN
 - Rosette形態素解析システム
 - Google, Adobe, Yahoo!, MSNなどの大手企業で導入

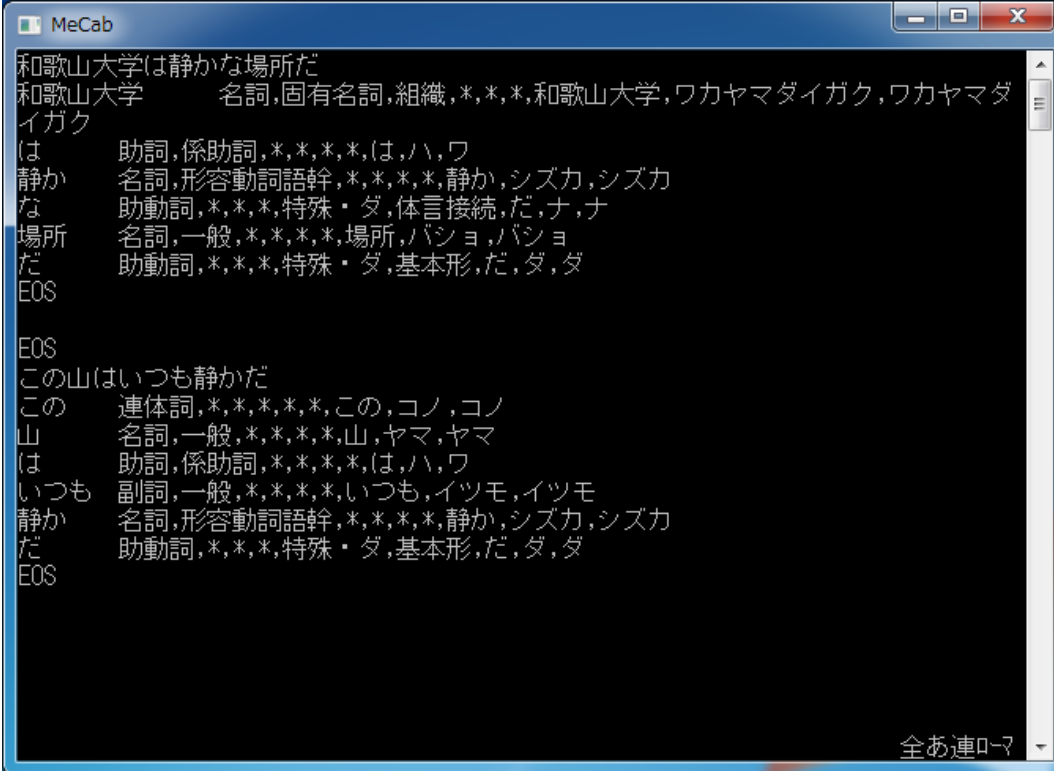
MeCabを使った解析例

和歌山大学は静かな場所だ

→和歌山大学 / は / 静か / な / 場所 / だ

この山はいつも静かだ

→この / 山 / は / いつも / 静か / だ



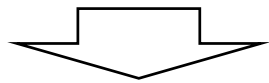
```
MeCab
和歌山大学は静かな場所だ
和歌山大学 名詞,固有名詞,組織,*,*,*,和歌山大学,ワカヤマダイガク,ワカヤマダ
イガク
は 助詞,係助詞,*,*,*,*,は,ハ,ワ
静か 名詞,形容動詞語幹,*,*,*,*,静か,シズカ,シズカ
な 助動詞,*,*,*,特殊・ダ,体言接続,だ,ナ,ナ
場所 名詞,一般,*,*,*,*,場所,バシヨ,バシヨ
だ 助動詞,*,*,*,特殊・ダ,基本形,だ,ダ,ダ
EOS

EOS
この山はいつも静かだ
この 連体詞,*,*,*,*,*,この,コノ,コノ
山 名詞,一般,*,*,*,*,山,ヤマ,ヤマ
は 助詞,係助詞,*,*,*,*,は,ハ,ワ
いつも 副詞,一般,*,*,*,*,いつも,イツモ,イツモ
静か 名詞,形容動詞語幹,*,*,*,*,静か,シズカ,シズカ
だ 助動詞,*,*,*,特殊・ダ,基本形,だ,ダ,ダ
EOS

全あ連0-7
```

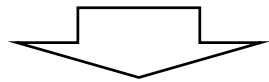
形態素解析によるインデックス作成

- 和歌山大学は静かな場所だ
- この山はいつも静かだ



形態素解析

- 和歌山大学 / は / 静か / な / 場所 / だ
- この / 山 / は / いつも / 静か / だ



重要な語をインデックス登録

- 和歌山大学, 静か, 場所
- 山, 静か

(余談) 学校文法との比較

- 学校文法: 学校で国語教育の一環として教える日本語文法のこと
- 学校文法では, 品詞を10種類に分類している
 - 名詞, 動詞, 形容詞, 形容動詞, 副詞, 連体詞, 感動詞, 接続詞, 助詞, 助動詞
 - 形容動詞の扱いについて, 学校文法以外では諸説紛々
- 形態素解析では, 品詞は形容動詞を除いた9種類と分類することが多い?
 - 形容動詞: 文中において単独で述語になることができ, 「～だ」で言い切れるもの
 - ChaSen, MeCabは(形容動詞) = (名詞) + (助動詞)
 - JUMANでは形容動詞は形容詞の扱い

まとめ

- 形態素解析は文を意味を持つ最小単位に分ける処理
- 人が使っている言語をコンピュータに処理させるための技術の1つ
- 入力から出力, データのやり取りで幅広く使用されている

参考文献

- 日本語形態素解析入門 Version 0.9.1 / 山下達雄
<<http://nais.to/~yto/doc/tech/jma/jma19990514.pdf>>
- 機械処理の日本語文法 / 金久保正明
<http://www.sist.ac.jp/~kanakubo/research/natural_language_processing/Japanese_grammer.html>
- 『日本語入力を支える技術』 / 徳永拓之
- 『入門 自然言語処理』 / Steven Bird, Ewan Klein, Edward Loper

東日本大震災

2014年4月25日

原田真帆

東日本大震災

- 2011年(平成23年)3月11日14時46分に発生
- 震源:宮城県沖
- マグニチュード:9.0、最大震度:7
- 死者:15885人、重軽傷者:6148人
- 特徴:波高10m以上の巨大津波が発生し、自然災害による経済損失額としては史上1位。

関東大震災

- 1923年(大正12年)9月1日11時58分に発生
- 震源:神奈川県沖
- マグニチュード:7.9
- 死者行方不明者:105000人余
- 特徴:液状化による地盤沈下、崖崩れ、沿岸部では津波による被害が発生し、日本災害史上最大級の被害。

阪神淡路大震災

- 1995年(平成7年)1月17日5時46分に発生
- 震源:兵庫県沖
- マグニチュード:7.3、最大震度:7
- 死者:6434人、重軽傷者:43792人
- 特徴:震度7が適用された初めての事例であり、戦後に発生した地震災害としては東日本大震災に次ぐ規模。

地震に関する情報通信分野の研究

- レジリエントな情報通信社会の実現のための産学官連携拠点(独立行政法人 情報通信研究機構)
- レジリエントな社会を実証するワイヤレステストベッドの構築(独立行政法人 情報通信研究機構)

※レジリエントな社会：災害が発生しても「つながる」ネットワーク、「壊れない」ネットワークを目指した、災害に強い情報通信技術を持つ社会のこと。

耐災害ICT研究センター

- 東日本大震災後に設立
→研究成果:「衛星通信」が誰でも「簡単に操作」を行うことができるよう解決。

NHK放送文化研究所 その1

- 調査対象:被災者

→災地でまず必要とされた情報は、個別具体的かつ局地的なローカル情報(居住地近辺の被害状況、個人の安否、避難所の場所、ライフラインの復旧予定、食料・水・ガソリンの入手方法、店舗の開店状況など)であったが、NHKや在京キー局が提供した被災地全体の状況や原発関連情報などはマクロな情報であり、ユーザーの緊急な情報ニーズを十分には満たしていなかった。

NHK放送文化研究所 その2

- 調査対象：被災者でない方
→ 家族・友人の安否確認など“パーソナルな情報”、交通情報や輪番停電など“ローカルな情報”、原発に関する“専門的な情報”などを取得。

東日本大震災による通信網への被害

国民生活上の重要なライフラインである、情報通信インフラにも甚大な被害が発生した。通信網については、東北・関東地方を中心に、回線の途絶や、停電等により情報通信機器が使用できなくなるなどの被害が発生した。

このような中、民間事業者等により、情報通信インフラの早期復旧に向けた取組が行われるとともに、公衆電話の無料化、特設公衆電話の設置等の災害対応の対策が実施された。また、放送による災害情報の提供や、インターネットを活用したソーシャルメディア等の新たなメディアが、安否確認や被災者支援のために使われるなど、新たな取組みも数多く行われた。

災害後の通信手段確保に向けた取り組み その1

- 災害用伝言サービスの提供
→固定電話、携帯電話、インターネットによって、災害用伝言サービスの提供を行った。
- 特設公衆電話の設置等
→避難所等に特設公衆電話(無料)を約2,300台設置した。
- 衛星携帯電話の貸与
→地震の影響を受けにくい衛星通信が通信手段として大きな役割を果たした。

災害後の通信手段確保に向けた取り組み その2

- 移動通信機器の貸与

→約1,770台のMCA無線や簡易無線の移動通信機器を貸し出した。

- アマチュア無線の活用

→アマチュア無線は、金銭上の利益のためではなく、業務用としては使用することはできない。しかし、災害発生時又は発生するおそれがあり、有線通信が利用できない、又は著しく利用が困難な場合には、人命の救助、災害の救援、交通通信の確保又は秩序の維持のために非常通信を行うことができることとされている(電波法第52条第4号)。

緊急地震速報

緊急地震速報とは、地震の発生直後に、震源に近い地震計でとらえた観測データを解析して震源や地震の規模(マグニチュード)を直ちに推定し、これに基づいて各地での主要動の到達時刻や震度を予想し、可能な限り素早く知らせる地震動の予報及び警報のことである。

気象庁では平成19年10月1日から、緊急地震速報の一般向け提供を行っている。

地上テレビ放送

地上テレビ放送については、親局への影響はなかったものの、東北、関東の被災各地における電力の途絶等により、中継局については東北6県を含む全11県で最大時120か所の停波が確認された。

このような中、NHK、地元放送事業者等は通常の番組編成を変更し、災害関連情報等の放送を行った。