

# 講義 (5)

## 「学術情報流通を実現する技術 (2) -応用技術」

佛教大学図書館 専門員

飯野勝則

平成25年度学術情報ウェブサービス担当者研修

2013年9月25日 at NII

# シンプルな学術情報流通

# 身近な例



CiNiiに大量の論文データを登録する、というのも学術情報流通の一形態

# CiNii (NII-ELS) に 大量のデータを登録する (1)

**NII-ELS**

一括登録

[設定変更](#)

[雑誌一覧](#) | [記事検索](#) | [一括登録](#) | [データ出力](#)

登録ファイルの形式	<input checked="" type="radio"/> TSV形式 <input type="radio"/> タグ形式
登録ファイル	<input type="button" value="参照..."/> ファイルが選択されていません。
TSV形式の種類 ※登録ファイルの形式がTSV形式の場合にのみ指定	<input checked="" type="checkbox"/> ダブルクォーテーションマークをくり文字とみなす

National Institute of Informatics

# TSV (Tab Separated Value) 形式

"AA12387978"	"38"	"20100301"	"P"	"非大学進学者のメンタリティに関する研究：三重県の分析を中心に"
"AA12387978"	"38"	"20100301"	"P"	"高齢者の自己概念の研究" "コウレイシャ ノ ジコ ガイネ"
"AA12387978"	"38"	"20100301"	"P"	"芥川龍之介小説にみる生の欲動と死の欲動の葛藤" "アクタガワ リ"
"AA12387978"	"38"	"20100301"	"P"	"「いのちの電話相談」の特質に関する実証的研究：修正版グラウン"
"AA12387978"	"38"	"20100301"	"P"	"児童期の家族関係と両親イメージが現在の自尊感情に与える影響"
"AA12387978"	"38"	"20100301"	"S"	"編集後記" "ヘンシュウ コウキ"
"AA12387934"	"38"	"20100301"	"P"	"次世代育成支援地域行動計画と保育所：地域子育て支援事業に見る理"
"AA12387934"	"38"	"20100301"	"P"	"韓国・老人長期療養保険制定の過程分析" "カンコク ロウジン チ"
"AA12387934"	"38"	"20100301"	"P"	"「荒れ」からの回復における虚構の遊びの可能性：個々の幼児と保"
"AA12387934"	"38"	"20100301"	"P"	"戦後の三池炭鉱における労務管理と労働者の抵抗に関する研究：三"
"AA12387934"	"38"	"20100301"	"P"	"地域施設通所精神障害者の身体的健康の課題" "チイキ シセツ"
"AA12387934"	"38"	"20100301"	"P"	"DV被害者の生活問題とソーシャルワークの課題：民間シェルターの"
"AA12387934"	"38"	"20100301"	"P"	"ライフ・サイクルによる貧困とワークハウス収容者：19世紀初期の"
"AA12387934"	"38"	"20100301"	"S"	"編集後記" "ヘンシュウ コウキ"
"AN10404998"	"95"	"20110301"	"P"	"現代語版『小説神髓』(一)" "ゲンダイゴ バン ショウセツ"
"AN10404998"	"96"	"20120301"	"P"	"現代語版『小説神髓』(二)" "ゲンダイゴバン ショウヤツ シ"

データ項目をタブによって切り分けたテーブルを連想させるフォーマット  
 簡単なデータの流通フォーマットとしてcsvと並び古くから利用されている  
 人の目では判別しにくい部分もある

# タグ形式

```

@LANG JPN↓
@YEAR 20110325↓
@AUTH 黒田,彰 / 坪井,直子 / 筒井大祐↓
@AUTY クロダ,アキラ / ツボイ,ナオコ / ツツイ,ダイスケ↓
@AUTE ↓
@ATTR P↓
@TITL 東原本八幡大菩薩後縁起 (上巻) : 影印, 翻刻↓
@TITY [ヒガシハラ] ハチマン ダイボサツ エンギ ジョウカン エイイン ホンコク↓
@TITE ↓
@VOLN 17↓
@@C KG00170010↓
@PAGE 5-34↓
@URL http://archives.bukkyo-u.ac.jp/infolib/meta\_pub/rid\_KG00170010↓
@ACCN ↓
@SRTN ↓
@RTYP ↓
@RTYE ↓
@FLNM ↓
@NCID AN10591104↓
@@C 更新レコードNo.2011/06/18-1↓
@@E ↓
@LANG JPN↓
@YEAR 20110325↓
@AUTH 浅尾,広良↓
@AUTY アサオ,ヒロヨシ↓
@AUTE ↓
@ATTR P↓
@TITL 皇女の婚姻から見た『源氏物語』 (中古特集) ↓
@TITY コウジョ ノ コンイン カラ ミタ ゲンジ モノガタリ チュウコ トクシュウ↓
@TITE ↓
@VOLN 17↓
@@C KG00170020↓
@PAGE 35-46↓
@URL http://archives.bukkyo-u.ac.jp/infolib/meta\_pub/rid\_KG00170020↓
@ACCN ↓

```

各行の先頭に「タグ」と呼ばれるメタデータ項目を判別するための目印をいれ、情報を伝達している。このような形での「タグ」形式はあまり一般的でない

# NII-ELSで採用された情報流通方式

- 人の手を介したプリミティブな学術情報流通の技術
- TSVはCSVとともに長い歴史があり、現在でも広く受容
- シンプルであるという点にフォーマットとしての価値
- 複雑なメタデータの場合には利用が困難

# 少し高度な学術情報流通 (自動化の技術)



# CiNii (NII-ELS) に 大量のデータを登録する (2)

## IRDBハーベスト概念図

学術機関リポジトリデータベース (IRDB) は OAI-PMH によって、日本の IR のメタデータを収集しています。IRDB に収集したメタデータは、JAIRO、CiNii 等で検索可能です。収集したメタデータは、IRDB コンテンツ分析システムで集計もしています。



①OAI-PMHに対応した②APIを用いて、IRDBに③XMLデータを収集させる

# ①OAI-PMH

(Open Archives Initiative - Protocol for Metadata Harvesting)

- Open Archives Initiativeにより策定されたメタデータを流通（交換）させるためのプロトコル（規格）
- XML形式で作成されたメタデータをHTTPプロトコルを用いて流通させる
- ウェブスケールディスクバリエーションなどが、自動で他のデータベースのメタデータを収集する「ハーベスト」作業の際などに標準的に用いられる



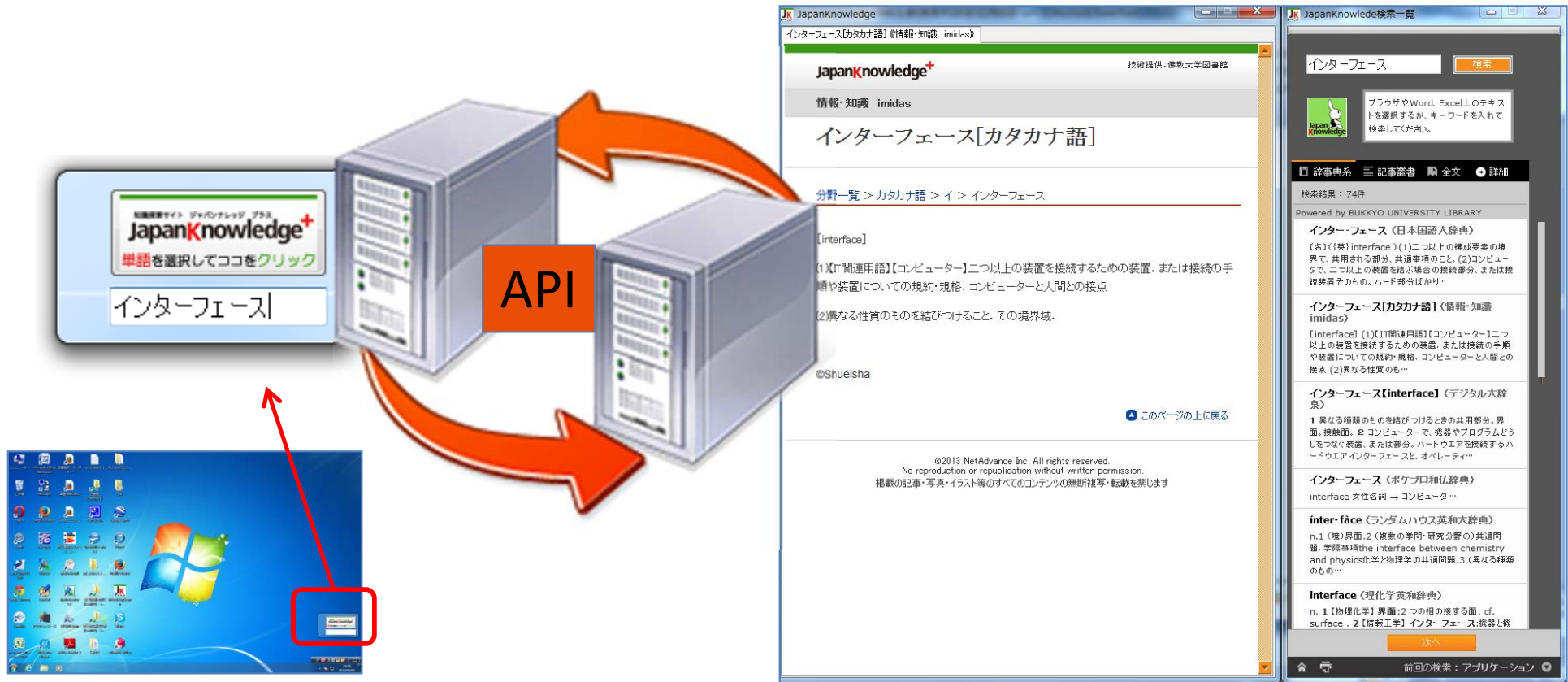
## ②API

# (Application Programming Interface)

- あるソフトウェアが外部のソフトウェアに対して提供するインターフェース（接続規格）
- データの外部からの呼び出しや、交換といった用途で用いられる
- 特にHTTPプロトコルを用いて、ウェブ上で用いられるAPIをウェブAPIと呼ぶ

# APIの使用例

## (デスクトップアプリ⇔データベース)



Windows用に作った常駐型アプリケーションの検索窓から「ジャパンナレッジ」を検索し、結果をオリジナルのブラウザで表示させる

## ③XML (eXtensible Markup Language)

- インターネット上のコンテンツを記述するための「データ記述言語」（⇔「プログラミング言語」）であり、タグによって各項目が「マークアップ」されている「マークアップ言語」のひとつ
- 1998年にW3C（World Wide Web Consortium）によりXML1.0が策定・勧告され、その後JIS規格化される。2004年にはXML1.1が策定された
- HTML（HyperText Markup Language）と異なり、タグを独自に定義できるという特徴がある
- 階層構造や親子化などの表現が容易であり、データの表示などに利用しやすい
- 「要素」と「属性」の使い方によって、同一のメタデータであっても、記述方法が変わるという特徴がある。どう使うか判断が難しい

# 「属性」と「要素」

```
<?xml version="1.0" ?>
<contacts>
  <contact name="田中 一郎" age="20" tel="1234-45△△" />
  <contact name="山田 花子" age="64" tel="4567-89△△" />
</contacts>
```

```
<?xml version="1.0" ?>
<contacts>
  <contact age="20">
    <name> 田中 一郎</name>
    <tel> 1234-45△△</tel>
  </contact>
  <contact age="64">
    <name>山田 花子</name>
    <tel> 4567-89△△</tel>
  </contact>
</contacts>
```

```
<?xml version="1.0" ?>
<contacts>
  <contact>
    <name>田中 一郎</name>
    <age>20</age>
    <tel> 1234-45△△</tel>
  </contact>
  <contact>
    <name>山田 花子</name>
    <tel> 4567-89△△</tel>
    <age> 64 </age>
  </contact>
</contacts>
```

適切か否かの判断を抜きにすれば「属性」「要素」の組み合わせにより、いくつかの記述を行うことができってしまう

実際にはタグや属性など構造を規定する「スキーマ」という定義により、データ流通が円滑に行われるように工夫されている

# XMLの使用例 (For OAI-PMH)

```

- <OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2013-08-14T08:05:07Z</responseDate>
  - <request metadataPrefix="junii2" verb="ListRecords" set="G0000019ronshu">
    http://archives.bukkyo-u.ac.jp/il4/oai_repository/pmh/G0000019-repository
  </request>
  - <ListRecords>
    - <record>
      - <header>
        <identifier>oai:infocom.co.jp:G0000019ronshu:O5003800001907</identifier>
        <datestamp>2013-07-31T15:01:20Z</datestamp>
        <setSpec>G0000019ronshu</setSpec>
      </header>
      - <metadata>
        - <junii2 xsi:schemaLocation="http://irdb.nii.ac.jp/oai http://irdb.nii.ac.jp/oai/junii2.xsd">
          <title>近藤治博士著作目録</title>
          <alternative>Bibliography of the Works of Dr. Osamu Kondo</alternative>
          <publisher>鷹陵史学会</publisher>
          <date>2012-12-11</date>
          <NIItpe>Others</NIItpe>
          <format>近藤治博士著作目録</format>
          <identifier>O5003800001907</identifier>
        - <URI>
          http://archives.bukkyo-u.ac.jp/repository/baker/rid_O5003800001907
        </URI>
        <issn>0386331X</issn>
        <NCID>AN00026748</NCID>
        <jtitle>鷹陵史学</jtitle>
        <volume>38</volume>
        <spage>1</spage>
        <epage>15</epage>
        <dateofissued>2012-09-29</dateofissued>
        <language>jpn</language>
      </junii2>
    </metadata>
  </record>
</ListRecords>
</OAI-PMH>

```

大学のリポジトリに登録したデータを出し、JAIROをはじめとする外部のデータベースやリポジトリにハーベストしてもらおう

# ユーザインタフェース上での 学術情報流通



# ウェブサービス間での連携例

The screenshot shows the Bukkyo University Library portal with several key features highlighted by red boxes and numbered 1 through 5:

- 1:** The search bar in the main navigation area.
- 2:** The RSS feed icon and the 'お知らせ' (Notice) section.
- 3:** The '佛教大学図書館デジタルコレクション' (Bukkyo University Library Digital Collection) section.
- 4:** The '佛教大学論文目録リポジトリ' (Bukkyo University Theses Bibliography) section.
- 5:** The footer area containing the '知識探索サイト JapanKnowledge' and other search services.

① SummonがAjaxを使ってキーワード候補を表示する

② RSSを使って別のウェブサイトと連携する

③ スクレイピングとJSONとAjaxを使って、「佛教大学デジタルコレクション」というアーカイブと連携する

④ APIとAjaxを使って「佛教大学論文目録リポジトリ」という機関リポジトリと連携する

⑤ APIを使って「ジャパンナレッジ」と連携する

# ① Ajax

## (Asynchronous JavaScript and XML)

- 非同期通信により、ウェブページの一部のみを更新する技術
- すなわち、ウェブページの再読み込みを行わずに、リアルタイムでその一部を書き換えることができる
- JavaScriptのAPIであるXMLHttpRequestを利用している

# Ajaxの例



佛教大学図書館 ポータルサイト  
BUKKYO UNIVERSITY LIBRARY

学術情報の収集や論文・レポート作成のための図書館サイト

お気軽検索 蔵書(OPAC) 雑誌ポータル データベース

利用案内 詳細検索 ヘルプ

検索

『蔵書(OPAC)』『佛教大学論文目録リポジトリ』などの学内コンテンツだけでなく、図書館が契約している電子ジャーナル、データベースなどの学術情報をシンプルに検索することができます。



佛教大学図書館 ポータルサイト  
BUKKYO UNIVERSITY LIBRARY

学術情報の収集や論文・レポート作成のための図書館サイト

お気軽検索 蔵書(OPAC) 雑誌ポータル データベース

利用案内 詳細検索 ヘルプ

京都 検索

Select an item from the list:

- 京都語文
- 京都議定書
- 京都の歴史
- 京都新聞

リポジトリ』などの学内コンテンツだけでなく、図書館が契約している電子ジャーナル、データベースなどの学術情報をシンプル

検索窓に入れた単語を使って、リアルタイムで検索候補を表示する

## ②RSS

- XMLをベースに策定された
- ウェブサイトの更新内容、要約、見出しなどを自動的に配信することを目的として使用される
- フォーマットや互換性などが異なる複数のバージョンが乱立している

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" media="screen" href="/d/styles/rss2full.xsl"?><?xml-stylesheet type="text/css" media="screen" href="http://feeds.
xmlns:content="http://purl.org/rss/1.0/modules/content/" xmlns:fw="http://wellformedweb.org/CommentAPI/" xmlns:dc="http://purl.org/dc/elements/1.1
xmlns:sy="http://purl.org/rss/1.0/modules/syndication/" xmlns:slash="http://purl.org/rss/1.0/modules/slash/" xmlns:feedburner="http://r

<channel>
  <title>佛敎大学図書館ポータルサイト</title>
  <link>http://bulib.bukkyo-u.ac.jp/docs/portal</link>
  <description>佛敎大学図書館ポータルサイトへようこそ。学術情報の収集や論文・レポート作成のための図書館サイト。</description>
  <lastBuildDate>Tue, 13 Aug 2013 00:27:02 +0000</lastBuildDate>
  <language>ja</language>
  <sy:updatePeriod>hourly</sy:updatePeriod>
  <sy:updateFrequency>1</sy:updateFrequency>
  <generator>http://wordpress.org/?v=3.0.4</generator>
  <atom10:link xmlns:atom10="http://www.w3.org/2005/Atom" rel="self" type="application/rss+xml" href="http://feeds.feedburner.com/ac/
  <atom10:link xmlns:atom10="http://www.w3.org/2005/Atom" rel="hub" href="http://pubsubhubbub.appspot.com/" /></atom10:link>
  <title>「新訂増補 国史大系」全巻が電子ブックとして利用できるようになりました</title>
  <link>http://feedproxy.google.com/~r/ac/Mx0i/3/u72mU0BVh90</link>
  <comments>http://bulib.bukkyo-u.ac.jp/docs/portal/?p=1270#comments</comments>
  <pubDate>Tue, 13 Aug 2013 00:27:02 +0000</pubDate>
  <dc:creator>佛敎大学 図書館</dc:creator>
  <category><![CDATA[新刊登録]]></category>
  <guid isPermaLink="false">http://bulib.bukkyo-u.ac.jp/docs/portal/?p=1270</guid>
  <description><![CDATA[日本史を研究するために必要な史籍の叢書である「新訂増補 国史大系」全6巻が電子ブックとして利用できるよ
u.ac.jp/docs/portal/?p=1270"><span class="meta-nav">[続き]</span></a>]]></description>
  <content:encoded><![CDATA[<p>日本史を研究するために必要な史籍の叢書である「新訂増補 国史大系」全6巻が電子ブックとして利用
</p><a href="http://archives.bukkyo-u.ac.jp/i14/meta_pub/G00000130BofDB_1935">NetLibrary eBook Collection</a>よりご利用ください。</p>
</ul>
  <li>国史大系 1 上日本書紀 (前) </li>
  <li>国史大系 1 下日本書紀 (後) </li>
  <li>国史大系 2 続日本書紀</li>
```

# RSSの例



**お知らせ**

2013年8月13日 **new** **新規登録**  
「新訂増補 国史大系」全巻が電子ブックとして利用できるようになりました

2013年8月2日 **重要**  
「佛教大学図書館デジタルコレクション」一部コンテンツの不具合について

2013年8月1日 **重要**  
「メディカルオンライン」での大量ダウンロード検出について

2013年7月23日 **重要** **メンテナンス**  
「浄土教報」「教学週報」などがサービス一時休止

2013年7月17日  
VPN接続サービスがWindows 8 / IE10に対応



佛教大学  
BUKKYO UNIVERSITY

アクセス・キャンパス一覧 ▶ サイトマップ ▶ 佛教大学教員検索 ▶ お問い合わせ

Language ▶ 資料請求 ▶

文字: 中 大 特 ▶ 検索

▶ 受験生の方へ ▶ 在学生の方へ ▶ 卒業生の方へ ▶ 保護者の方へ ▶ 進路指導担当の方へ ▶ 企業の方へ ▶ 一般の方へ

大学紹介 ▶ 学部・大学院 ▶ 入試情報 ▶ 学生生活 ▶ キャリア・就職 ▶ 図書館・附置機関 ▶ 研究・社会連携 ▶ 通信教育

TOP > 図書館・附置機関 > 図書館

図書館  
Bukkyo University Library

大学の重要な学術情報基盤として、学生の学習や教員の研究・教育活動を積極的に支援しています。

図書館ポータルサイトからの最新のお知らせ

2013年8月13日 ▶ 「新訂増補 国史大系」全巻が電子ブックとして利用できるようになりました

2013年8月2日 ▶ 「佛教大学図書館デジタルコレクション」一部コンテンツの不具合について

2013年8月1日 ▶ 「メディカルオンライン」での大量ダウンロード検出について

2013年7月23日 ▶ 「浄土教報」「教学週報」などがサービス一時休止

2013年7月17日 ▶ VPN接続サービスがWindows 8 / IE10に対応

図書館・附置機関

図書館  
Bukkyo University Library

- ▶ 図書館を利用される方へ
- ▶ 図書館利用規程
- ▶ 図書館データ
- ▶ 紫野キャンパス図書館フロアマップ
- ▶ 二条キャンパス図書室
- ▶ 図書館広報誌
- ▶ 浄土宗文献室
- ▶ お問い合わせ

ポータルサイトに登録した情報を別サイトで同期して表示させる

## ③スクレイピング

- HTMLで作成されたウェブページを対象に、必要な部分だけ取得・整形して、自サイトにおいて利用する技術
- ウェブAPIがないウェブページなどの情報を取得する際に利用される

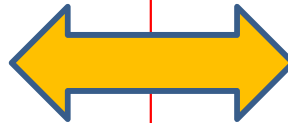
## ④JSON (JavaScript Object Notation)

- XMLと同様の「データ記述言語」
- XMLよりも小さく軽量で、かつJavaScriptへの親和性が高いため、ajaxの普及とともに利用が盛んになった
- 「スキーマ」という概念はもともとはなかったが、JSONの普及が進んだことから、近年その策定が進んでいる

# XMLとJSON

```
<?xml version="1.0" ?>
<contacts>
  <contact>
    <name>田中 一郎</name>
    <age>20</age>
    <tel> 1234-45△△</tel>
  </contact>
  <contact>
    <name>山田 花子</name>
    <age> 64 </age>
    <tel> 4567-89△△</tel>
  </contact>
</contacts>
```

XML      JSON



```
{
  "contacts": {
    "contact": [
      {
        "name": "田中 一郎",
        "age": "20",
        "tel": "1234-45△△"
      },
      {
        "name": "山田 花子",
        "age": "64",
        "tel": "4567-89△△"
      }
    ]
  }
}
```

```
<?xml version="1.0" ?>
<contacts>
  <contact name="田中 一郎" age="20" tel="1234-45△△" />
  <contact name="山田 花子" age="64" tel="4567-89△△" />
</contacts>
```



```
{
  "contacts": {
    "contact": [
      {
        "-name": "田中 一郎",
        "-age": "20",
        "-tel": "1234-45△△"
      },
      {
        "-name": "山田 花子",
        "-age": "64",
        "-tel": "4567-89△△"
      }
    ]
  }
}
```



# スクレイピングとJSONの例

①アーカイブサイト  
「佛教大学図書館デジタルコレクション」

④ポータルサイトにて、  
ajaxを使って表示



```

"dl": {
  "dt": {
    "span": {
      "class": "icon clctn",
      "content": "コレクション"
    },
    "a": {
      "href": "http://archives.bukkyo-u.ac.jp/collections/?collection=senchakuhongankencho",
      "content": "選擇本願念佛集 建長版"
    }
  },
  "dd": {
    "p": {
      "a": {
        "class": "more-link",
        "href": "http://archives.bukkyo-u.ac.jp/collections/?collection=senchakuhongankencho",
        "content": "もっと読む"
      },
      "content": "建長頃刊 半面六行十七字 縦二五・五種、横一五種 鳥子 粘葉装 修復あり 浄土宗宗室 無刊記 [本末見返し墨 ...
    }
  }
},
{
  "class": "box-frag01 tile03 boxLink",
  "diy": {
    "class": "img-outside img-transparent",
    "a": {

```

②HTMLをJSONとして取得

```

デジタルコレクション
DigitalCollectionImg = function(){
  $("#digital-collections-images").html("<img src='/docs/portal/wp/
$.getJSON("http://query.yahooapis.com/v1/public/yql?callback=?",
    q: "select * from html where url='http://archives.bukkyo-u.ac.jp/collections/?collection=senchakuhongankencho'
    format: "json"
  ), function (d) {
    var i = Math.floor( Math.random() * 5 );
    var items = d.query.results.digit
    num = te
  }

  if(!num){
    $("#digital-collections").hide();
    return false;
  }

  var img = items[i].a.img.src,
      url = items[i].a.href,
      title = items[i].dl.dt.a.content,
      note = items[i].dl.dd.p.content,
      html = "<figure><a href='"+url+"' target='_blank'><img src
$("#digital-collections-images").html(html);

```

③JavaScript (jQuery) を  
使ってスクレイピング

# 「学術」を超えて

# 図書館員と情報技術

- 最新の情報技術は図書館が開発したものではない
- 最新の情報技術は「学術」のためだけに限定されるものではない
- 学術情報を扱う際に必要とされる技術は閉じたものではなく、ウェブ世界との密接なつながりのなかにある
- 図書館員は最新の情報技術を知り、教育や研究のために何ができるのかを常に考える必要があるのでは？

ご清聴ありがとうございました