

IUGONETシステム報告

阿部修司[1], 梅村宜生[2], 小山幸伸[3], 堀智昭[2], 谷田貝亜紀代[4], 新堀淳樹[4], 田中良昌[5], 上野悟[6], 金田直樹[6],
八木学[7], 佐藤由佳[5], IUGONETプロジェクトチーム

[1] 九州大・国際宇宙天気科学・教育センター、

[2] 名古屋大・太陽地球環境研究所、

[3] 京都大・理・地磁気センター、[4] 京都大・生存圏研究所、

[5] 国立極地研究所、[6] 京都大・理・附属天文台、

[7] 東北大・惑星プラズマ大気研究センター

IUGONET機関の所有する観測機器

アイスランド
オーロラ観測(2点)
地磁気観測(3点)

トロムソ
ISレーダー
流星レーダー
MFLレーダー

信楽MU
観測所

北海道HFレー
ダー(探査範囲)

太陽望遠鏡

赤道大気レーダー(EAR)

飯舘・女川観測所

スバルバル: ISレーダー、
流星レーダー、オーロラ観測



昭和基地
SuperDARNレーダー2台
MFLレーダー
オーロラ観測
地磁気多点観測点網
昭和・女川 ELF 同時観測



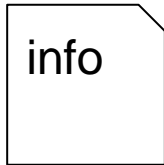
- MSTレーダー
- ▲ MF / 流星レーダー
- ◆ MAGDAS/地磁気観測
- ★ FM-CWレーダー
- OMTI光学観測
- WDC/地磁気観測所

- 自分の専門のデータ(例えば地磁気指数)はうまく使えるが、種類の違うデータ(例えば太陽撮像データ)を同時に使うための各情報入手することが難しく、総合解析になかなか結びつかない
- 他の分野のデータに関する情報がなかなか得られないストレス
- このような問題を解決する手段として、IUGONETは観測データのメタデータを共有するシステムと統合解析ツールを開発
- 本講演では、前者の観測データのメタデータ・データベース(IUGONET MDB)について紹介する

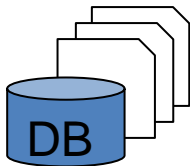


メタデータ・データベースの利用

- メタデータ・・・観測データそのものではなく、
観測データに付帯した情報
e.g. 観測開始終了時刻、観測場所、
観測データの所在情報
→ メタデータを介して
観測データへアクセス



- メタデータ・データベース・・・**メタデータを集めて**
検索可能にしたもの



様々な研究機関にある観測データのメタデータを、
1クエリーで検索可能

(いつ、どこで、どの物理量が同時観測されたのか?)



- フリーソフトウェア **DSpace** をベースに構築

- <http://www.dspace.org/>
- 学術機関リポジトリで広く使用されているため、情報の入手が容易
- 少々のカスタマイズ (JSP、Servlet) で IUGONET メタデータを利用可能
- 採用しているメタデータのベースは SPASE (<http://www.spase-group.org/>)
- Dspace では直接 XML を扱えないため、フォーマット変換して格納 (プログラムを IUGONET にて作成 = g2d, git2dspace)

- 2012年3月より **運用中**

- β版公開時 (2011年5月) はサーバ1台のみ (iugonet1.stelab) の運用
- 2012年1月にメインサーバを移行 (iugonet7.serc)。
- 旧メインサーバはサブとして使用し、現在、メインサブ2台の **冗長体制**



問題点、検討事項

1. 現行のメタデータ登録プログラム(g2d)の諸問題(e.g., 登録失敗時の明確な**手戻り手段**・多重登録の**確認手段がない**、など)
2. メタデータ増大に伴う各種パフォーマンス低下の問題=**登録から検索できるようになるまでに非常に時間がかかる**。**メタデータの置き換え、削除**処理を入れると所要時間が倍増する
3. 単一マシンでの運用限界=例えば、ベースになっている検索エンジン(Apache Lucene)において、**1000万件を超えると検索パフォーマンスが急激に低下する**恐れがある
4. **ユーザーが使いやすいサービスの提供**

などなど

3.2. 登録対象リポジトリの設定 (repoconf)

どのリポジトリを登録対象とするかを、repoconf コマンドを使って設定します。

3.2.1. 設定情報の一覧表示

repoconf コマンドの -a オプションを使います。(--all オプションでも実行できます。)

```

$ ./repoconf -a
2013/04/15 10:34:20.910573 ###[INFO] ##### Show Repository List (All) #####
Repository (Code: Name) | Activity | RemoteHost | RemotePath
-----
1: KyotoU_Obs | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_Observatory.git
2: KyotoU_Obs#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_Observatory/Granule.git
3: RISH1 | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_RISH.git
4: RISH1#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_RISH/Granule.git
...
23: JMA | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/JMA.git
24: JMA#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/JMA/Granule.git
    
```

初期状態では、すべてのリポジトリが Activity = 9: STOP (登録対象外) の状態に設定されています。

3.2.2. リポジトリを登録対象に変更する

repoconf コマンドの -active オプションを使います。

```

$ ./repoconf -active 1
    
```

• 新g2d (git2dspace) の開発

– パフォーマンスの改善

- アルゴリズムの全面見直し
- 約2.5倍の高速化

– 運用面の改善

- 不測の事態に備えたバックアップ、手戻り機能の実装
- 機関毎のリポジトリに対応。登録できるリポジトリ細分化
- 運用マニュアルの整備

– 現在βテスト中

- 分散型MDBでも使用中

```

[dspace@lunallena77 ~]$ repoconf -a
2013/08/17 16:10:19.818375 ###[INFO] ##### Show Repository List (All) #####
Repository (Code: Name) | Activity | RemoteHost | RemotePath
-----
1: KyotoU_Obs | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_Observatory.git
2: KyotoU_Obs#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_Observatory/Granule.git
3: RISH1 | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_RISH.git
4: RISH1#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_RISH/Granule.git
5: RISH2 | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_RISH_2.git
6: RISH2#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_RISH_2/Granule.git
7: SPEL | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_SPEL.git
8: SPEL#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_SPEL/Granule.git
9: WDC | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_WDC.git
10: WDC#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyotoU_WDC/Granule.git
11: KyushuU | 1: ACTIVE | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyushuU.git
12: KyushuU#Granule | 1: ACTIVE | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/KyushuU/Granule.git
13: NIPR | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/NIPR.git
14: NIPR#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/NIPR/Granule.git
15: STEL | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/STEL.git
16: STEL#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/STEL/Granule.git
17: TohokuU | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/TohokuU.git
18: TohokuU#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/TohokuU/Granule.git
19: NAOU | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/NAOU.git
20: NAOU#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/NAOU/Granule.git
21: NICT | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/NICT.git
22: NICT#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/NICT/Granule.git
23: JMA | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/JMA.git
24: JMA#Granule | 9: STOP | iugonet7.serc.kyushu-u.ac.jp | /git/git/Metadadata/Draft/JMA/Granule.git
    
```

2013/07/26	22:24:27
2013/07/26	22:24:27
2013/07/26	22:24:27
2013/07/26	22:24:27
2013/07/26	22:24:27
2013/07/26	22:24:27

遅い。。。

運用に
大きい
支障！！

メタ
データ登録



メタ
データ登録



メタデータ
800万件登録



メタデー
かか

メタデータを登録しても
MDBに反映されるまでに
時間がかかる



対応：MDBの分散化

- MDBを分散化することにより、現在のシステムを大きく変更することなく、**大幅なパフォーマンスアップ**が可能
- 検索システムの分散化は**大規模システムでは一般的**に行われている
 - 例えば、Googleでは1クエリに対し数千台のマシンで処理が行われている
- OpenSearchにより、メタデータのハーベスティングなしに**複数サーバの検索結果を表示可能**
 - 横断検索サーチエンジンの例：国立国会図書館サーチ (NDL Search)
 - 他プロジェクトとのメタデータ交換が容易に

OpenSearch: 検索結果を他のサイトから自由に利用できるようにするための仕組み

検索結果を他から利用する拡張に適應できるため、これを分散型MDBの構築に利用する

OpenSearch仕様は
<http://www.iugonet.org/opensearch.html>

http://search.iugonet.org/iugonet/opensearch/request?query=nipr_1sec_fmagg_syo_&ts=2010-01-01&te=2010-01-05&Granule=granule



**OpenSearchの結果
(ブラウザでの結果表示)**

分散化MDB

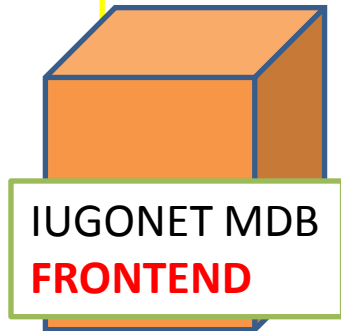


検索クエリ

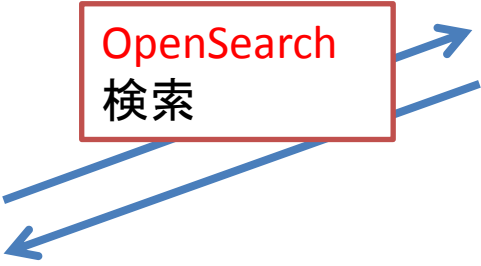


各サーバの結果をマージして返す

ユーザーは全サーバの情報を検索可能



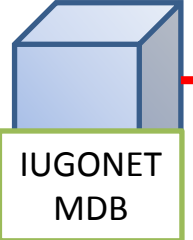
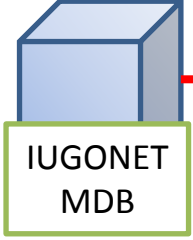
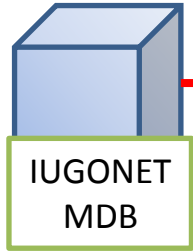
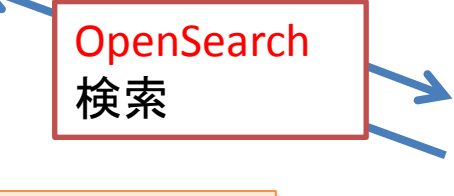
OpenSearch
検索



OpenSearch
検索

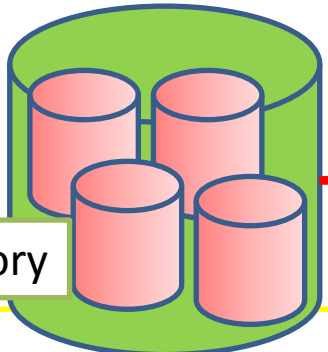


OpenSearch
検索



IUGONET MDBを分散して登録負荷を下げる

MD Repository



分散化MDB(テスト中)

HOST STATUS

- TOHOKU
- NIPR
- STEL
- RISH
- WDC
- HIDA_OBS
- ICSWSE

各サーバの運用状況が一覧可能

XML Interface

- ATOM
- RSS1
- RSS2

W3C XHTML 1.0

各サーバからの検索結果数一覧
特定サーバのみの結果表示も可能

Copyright (c) 2009-2013 IUGONET -Feedback
Customized by IUGONET.
Powered by DSpace Software.

[Standard radio wave phase-amplitude variation data](#) XML (via. TOHOKU)

Data Set
Amplitude and phase time series of low frequency standard radio transmitter signals at 19.80(NWC), 21.40(NPM), 24.80(NLK), 25.20(NDK), 37.50(NRK), 40.75(NAU), and 60.00kHz(WWU) ...
Start Date: 2010-10-24T00:00:00
Stop Date: -P2D
<http://iprt.gp.tohoku.ac.jp/lf/cdf/ath/>
Repository: spase://IUGONET/Repository/TohokuU/lf_repository
Instrument: spase://IUGONET/Instrument/TohokuU/radio_obs/ath/lf_ath

[Geomagnetic field data measured at ONAGAWA observatory, Tohoku University](#) XML (via. TOHOKU)

Data Set
Geomagnetic field data measured using Onagawa search coil magnetometer.
Start Date: 1991-01-01T00:00:00
Stop Date:
<http://adelie.geophys.tohoku.ac.jp/~onagawa/>
Repository:
Instrument: spase://IUGONET/Instrument/TohokuU/mag_obs/onw/sm

[Aerosol observation at Shigaraki in 2010](#) XML (via. STEL)

Data Set
Aerosol observation at the Shigaraki MU observatory of the Research Institute for Sustainable Humanosphere (RISH), Koka, Shiga, Japan
Start Date: 2010-10-27T10:00:00
Stop Date:
<http://www.stelab.nagoya-u.ac.jp/ste-www1/div1/>

結果には所属が表示される

BREAKDOWN

- TOHOKU(14)
- NIPR(70)
- STEL(65)
- RISH(17)
- HIDA_OBS(19)
- ICSWSE(115)

Results

IUGONET

Time From: YYYY-MM-DDThh:mm:ssZ to YYYY-MM-DDThh:mm:ssZ [UTC]

Data Set (Numerical Plot / Movie) Data File / Plot Instrument Observatory

Catalog Person Service
 Document Annotation Repository Registry

Search

Results 1-10 of 300

Results/Page 10 | Sort Items by Resource Type in Order Descending | Sort!

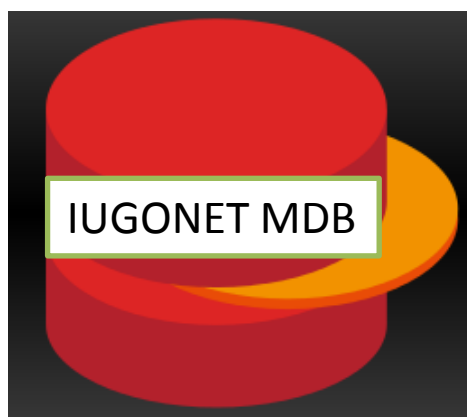
Result on XML

- ATOM
- RSS1
- RSS2

W3C XHTML 1.0

現在パフォーマンスチューニング中
今年度下半期に導入予定

- ・専門分野外のメタデータ検索をする場合、**検索語句の選択が困難**
 - ・専門に近い分野であっても、正しく検索語句を選べるとは限らない
= **メタデータに記載されている語句でなければヒットしない**
- 等の解決に役立つ



Substorm

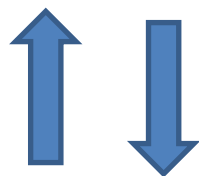


CME, AE,
Ionosphere...



国立情報学研究所で
開発された連想検索システム
GETAssoc の導入を検討

「Substorm」
を検索してみよう



メタデータヒットなし
関連用語は
CME, AE, ionosphere...



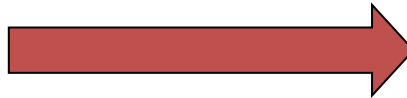
検索語句そのものがヒットしなくても、
関連する用語を検索語句候補として
表示することができる。

どのような用語が連想されるのかは、連想検索システムに登録している辞書次第
 Googleスプレッドシートで管理した語句テーブルを辞書ファイルに変換するしくみを
 開発→辞書の共有、複数人による辞書作成

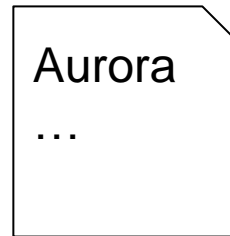
Google Spreadsheet

A	B
6	Thermosphere The thermosphere is the layer of the Earth's atmosphere directly above the mesosphere and directly below the exosphere.
7	Mesosphere The mesosphere is the layer of the Earth's atmosphere that is directly above the stratosphere and directly below the thermosphere.
8	Stratosphere The stratosphere is the second major layer of Earth's atmosphere, just above the troposphere, and below the mesosphere.
9	Troposphere The troposphere is the lowest portion of Earth's atmosphere. It contains approximately 80% of the atmosphere's mass and 99% of its water vapor and aerosols.
10	Disturbance storm time index The disturbance storm time (Dst, Kyoto Dst) index is a measure in the context of space weather.
11	K-index The K-index quantifies disturbances in the horizontal component of earth's magnetic field with an integer in the range 0-9 with 1 being calm and 9 or more indicating a geomagnetic storm.
12	Super Dual Auroral Radar Network The Super Dual Auroral Radar Network (SuperDARN) is an international radar network for studying the upper atmosphere and ionosphere, comprising seven radars in the northern hemisphere and seven in the southern hemisphere that operate in the High Frequency (HF) bands between 8.0 MHz (37m) and 22.0 MHz (14m).
13	Japan Meteorological Agency The Japan Meteorological Agency, frequently abbreviated to JMA, is an agency of Ministry of Land, Infrastructure, Transport and Tourism.
14	International Association of Geomagnetism and Aeronomy (IAGA) The International Association of Geomagnetism and Aeronomy (IAGA) is the premier international scientific association promoting the study of terrestrial and planetary magnetism and space physics.
15	Committee on Data for Science and Technology The Committee on Data for Science and Technology (CODATA) was established in 1966 as an interdisciplinary committee of the International Council for Science. It seeks to improve the compilation, critical evaluation, storage, and retrieval of data of importance to science and technology.
16	World Data Center The World Data Centre (WDC) system was created to archive and distribute data collected from the observational programmes of the 1957-1958 International Geophysical Year by the International Council of Science (ICSU).
17	International Council for Science (ICSU) The International Council for Science (ICSU) is an international non-governmental organization devoted to international cooperation in the advancement of science.
18	Intermagnet Intermagnet (International Real-time Magnetic Observatory Network) is a global network of observatories, monitoring the Earth's magnetic field.
19	CHAMP CHAllenging Minisatellite Payload (CHAMP) was a German satellite launched July 15, 2000 from Plesetsk, Russia and was used for atmospheric and ionospheric research, as well as other geoscientific applications, such as GPS radio occultation.
20	National Institute of Polar Research National Institute of Polar Research is the Japanese research institute for Antarctica. The agency manages several research bases on the continent.
21	International Geomagnetic Reference Field The International Geomagnetic Reference Field (IGRF) is a standard mathematical description of the Earth's main magnetic field.
22	International Reference Ionosphere The International Reference Ionosphere (IRI) is a common permanent scientific project of the Committee on Space Research and the International Union of Radio Science started 1960/65. It is the international standard for the terrestrial ionosphere since 1960.

語句テーブルのダウンロード
 辞書ファイルへの変換



辞書ファイル(.itb)



例:
 キーワード「substorm」での検索数0
 →「CME」「AE index」などが関連キーワード
 検索結果へのリンク作成、表示

Search Results

substorm
 Time from: [] to []
 Data Set (Numerical Plot / Movie) Data File / Plot

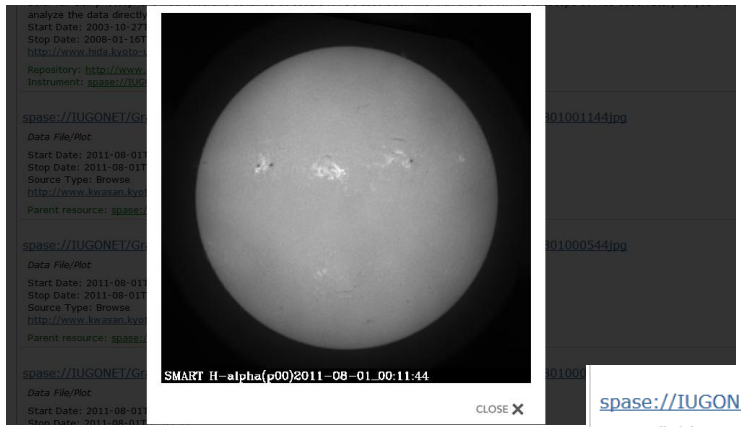
Search produced no results.

Results/Page 10 | Sort items by Relevance

Did you mean:

- [CME](#)
- [AE index](#)
- [Ionosphere](#)
- [Intergovernmental Panel on Climate Change](#)
- [Exosphere](#)

1. Lightbox2を用いたDisplay データの表示

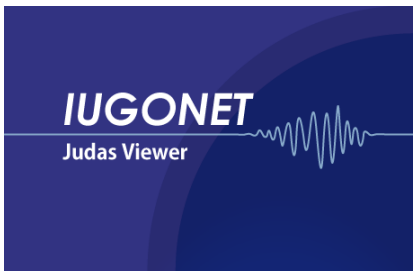
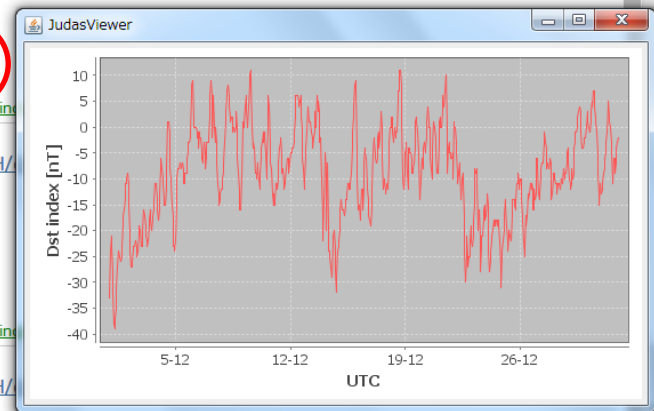


左記機能に加えて。。。
↓

[spase://IUGONET/Granule/WDC_Kyoto/WDC/Dst/index/PT1H/dst198612_wdc](http://wdc-data.iugonet.org/data/hour/index/dst/1986/dst8412)
Data File/Plot
Start Date: 1986-12-01
Stop Date: 1986-12-31
Source Type: Data
<http://wdc-data.iugonet.org/data/hour/index/dst/1986/dst8412>
Parent resource: [spase://IUGONET/NumericalData/WDC_Kyoto/WDC/Dst/index/PT1H/dst198612_wdc](http://wdc-data.iugonet.org/data/hour/index/dst/1986/dst8412)

[spase://IUGONET/Granule/WDC_Kyoto/WDC/Dst/index/PT1H/dst195812_wdc](http://wdc-data.iugonet.org/data/hour/index/dst/1958/dst5802)
Data File/Plot
Start Date: 1958-02-01
Stop Date: 1958-02-28
Source Type: Data
<http://wdc-data.iugonet.org/data/hour/index/dst/1958/dst5802>
Parent resource: [spase://IUGONET/NumericalData/WDC_Kyoto/WDC/Dst/index/PT1H/dst195812_wdc](http://wdc-data.iugonet.org/data/hour/index/dst/1958/dst5802)

[spase://IUGONET/Granule/WDC_Kyoto/WDC/Dst/index/PT1H/dst195812_wdc](http://wdc-data.iugonet.org/data/hour/index/dst/1958/dst5802)
Data File/Plot
Start Date: 1958-01-01
Stop Date: 1958-01-31
Source Type: Data



※JUDASの詳しい説明はソフトウェアの発表にて

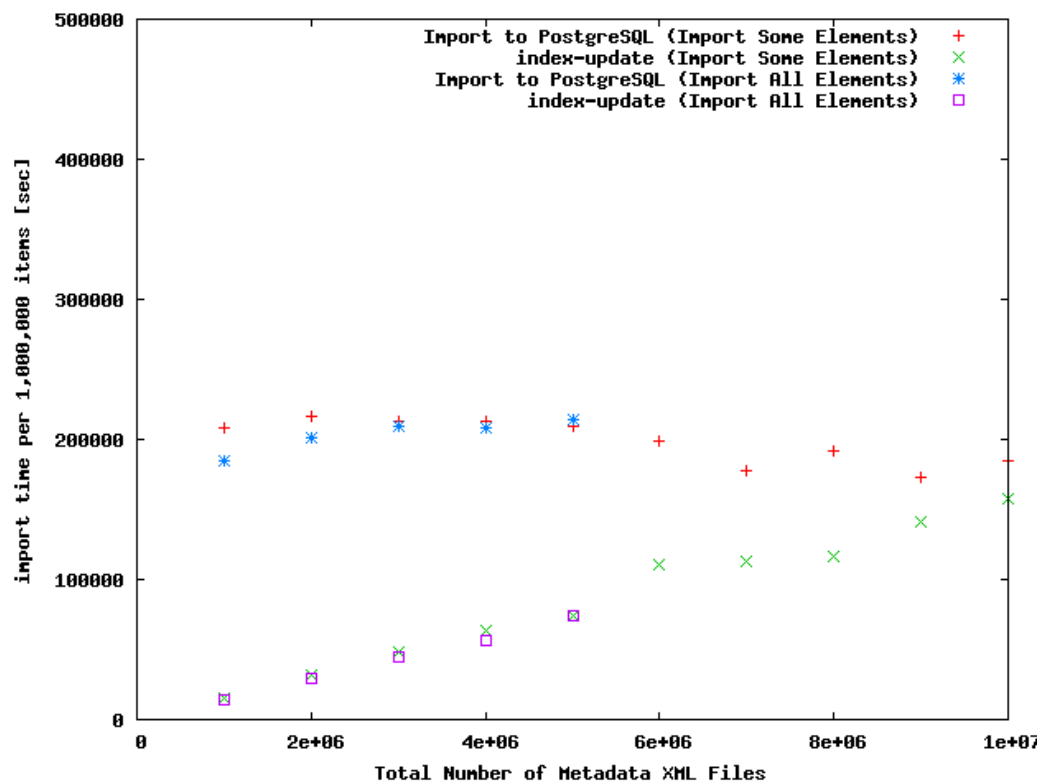
2. Java Web Start を用いたGranuleデータの表示

全SPASE要素の登録

これまで: 検索に必須と思われるSPASE要素を厳選して登録

今後: **全てのSPASE要素を登録**=全SPASE要素が検索対象

これまでキーワード検索ではヒットしなかったメタデータもヒットするようになる

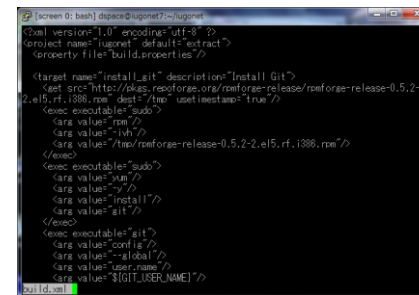


全要素登録に伴う登録時間パフォーマンス

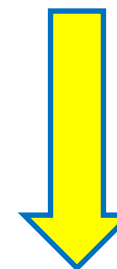
- 全要素を登録しても、PostgreSQL登録時間、Luceneインデキシング時間は一部要素の登録結果と変わらず、全要素登録に伴うパフォーマンス低下は発生しない。
- PostgreSQLのmetadatatypevalueテーブルのレコード数は倍増する。しかしながら、DSpaceはキー項目から値を取得しているため、検索パフォーマンスにも影響はないと考えられる。

build.xml

- **Github**(<https://github.com/iugonet>)にて、MDB含む多数の
プロダクトを準備
 - IUGONET MDB
 - 連想検索システム
 - JavaベースのUDAS(解析ソフトウェア)
 - etc etc...
- 利用された場合はIUGONET
開発員までフィードバックをい
ただけると助かります



```
<?xml version="1.0" encoding="UTF-8" ?>
<project name="iugonet" default="extract"
  <property file="build.properties" />
  <target name="install_git" description="Install Git"
    <get src="http://kias.repository.org/antforze-release-0.5.2-2.el5.rf.i386.rpm" dest="${tmp_dir}/antforze-release-0.5.2-2.el5.rf.i386.rpm" />
    <exec executable="sudo"
      <arg value="rm" />
      <arg value="-f" />
      <arg value="${tmp_dir}/antforze-release-0.5.2-2.el5.rf.i386.rpm" />
    </exec>
    <exec executable="sudo"
      <arg value="yum" />
      <arg value="-y" />
      <arg value="install" />
      <arg value="git" />
    </exec>
    <exec executable="git"
      <arg value="config" />
      <arg value="--global" />
      <arg value="user.name" />
      <arg value="${GIT_USER_NAME}" />
    </exec>
  </target>
</project>
```



1. Java
2. ant
3. build.xml



IUGONET PRODUCTS

- 公開中のメタデータ・データベース、及び各種ツールは順調に運用を続けている
- メタデータ数の劇的な増加による、各種パフォーマンスの低下に対応するため、メタデータ登録ソフトウェアの大幅な改良、MDBの分散化などを実施している
- 連想検索システムやQuicklookの導入など、ユーザーにとって便利な機能の実装を進めている
- 今後もより良いシステムの構築・提供を進めていく