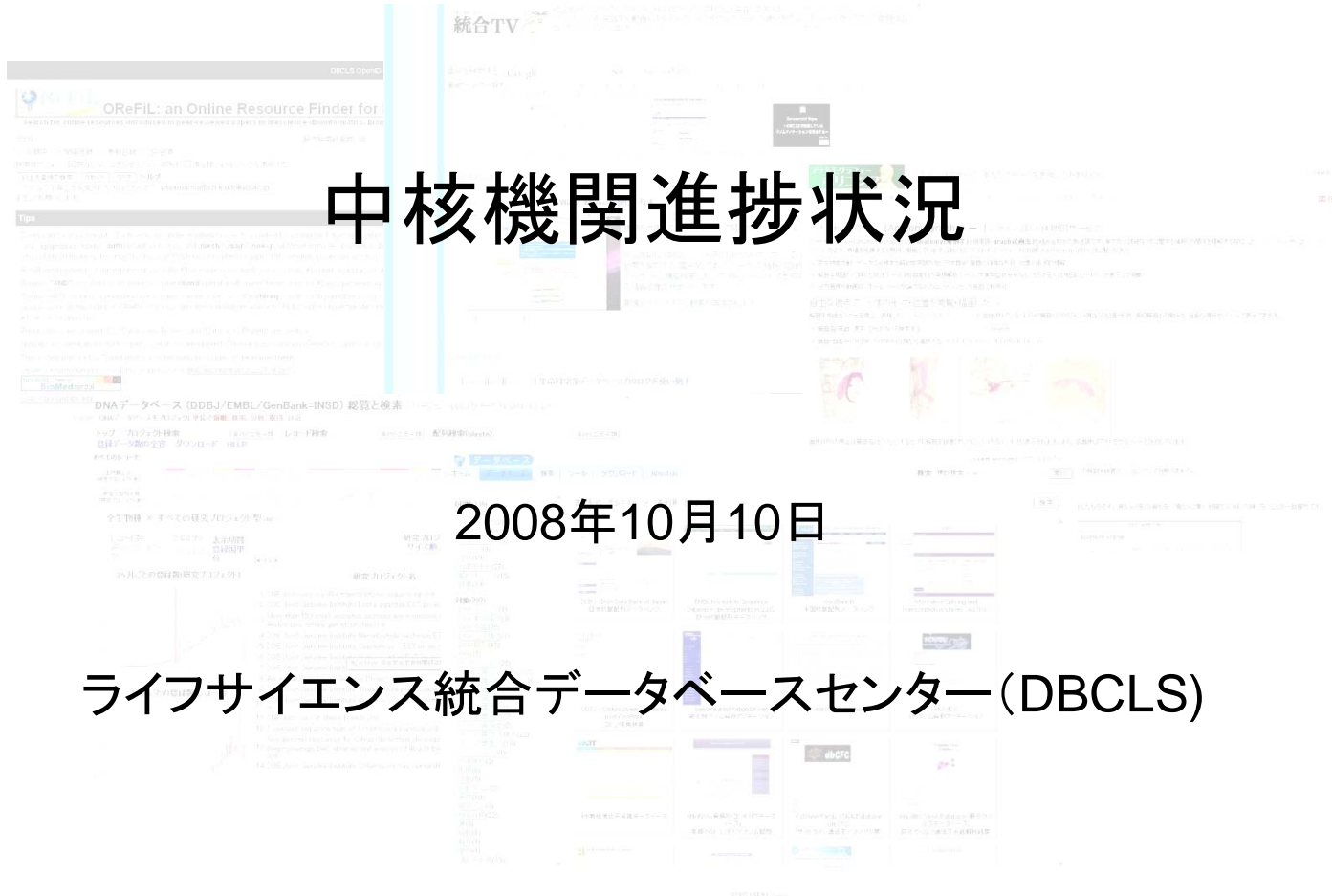


統合データベースプロジェクト 研究運営委員会作業部会資料



中核機関進捗状況

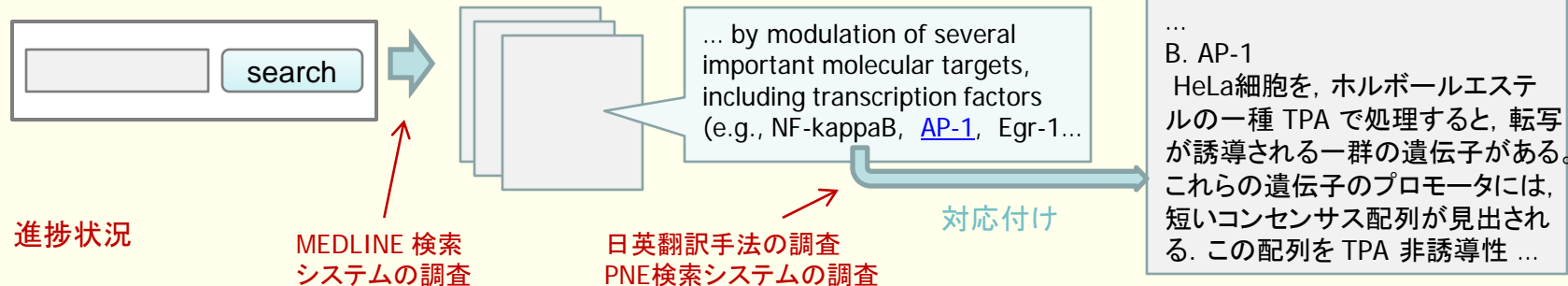
2008年10月10日

ライフサイエンス統合データベースセンター(DBCLS)

共通基盤技術開発

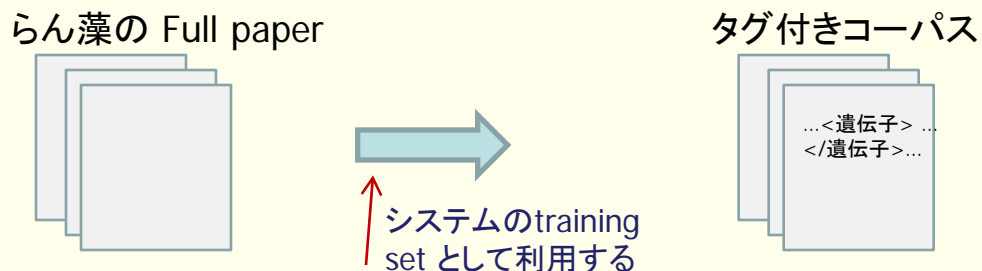
英文テキスト中の領域固有語を同定し、日本語に翻訳して、日本語文献を検索するシステムを開発

システム概念図



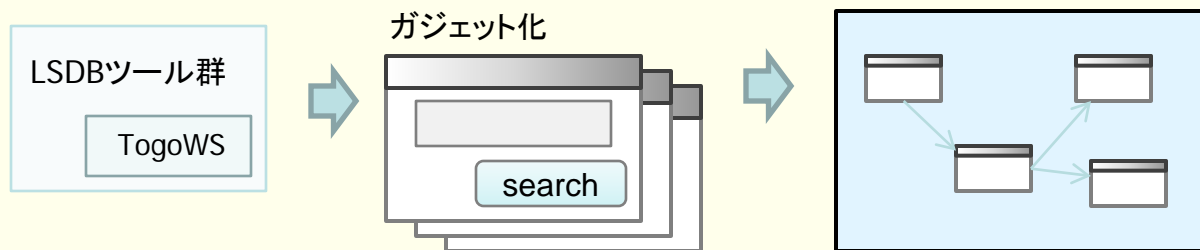
遺伝子名、蛋白質名で full paper を対応付ける技術を開発し、特定生物を対象としたプロトシステムを完成

システム概念図



TogoWSの操作性向上のためのツール組合せインターフェイス設計のための調査を行い、報告資料を作成

システム概念図



進捗状況

Yahoo pipes!, UCSC Galaxy など, 既存のツール組み合わせインターフェイスの調査

辞書の高度化とライフサイエンス統合検索

専門世界へ検索口

不明瞭な質問用語
p53, 肝癌

専門世界の目次

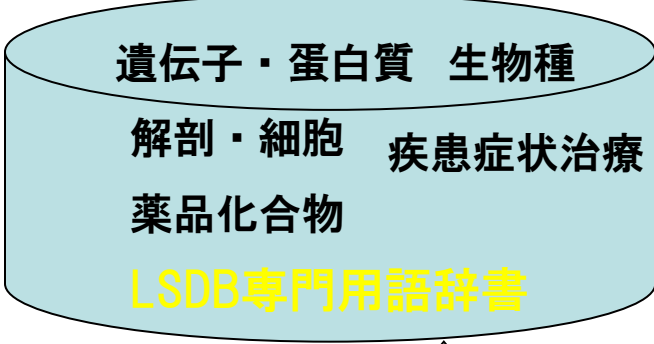
クエリ不在の興味

ステップ1. 「専門概念」を正しく特定してもらう助け

- 部分マッチする専門用語
- 上位、下位用語

Entrez Gene
 Entrez Med
 Entrez Pubmed
 Entrez Snp
 Entrez Tdb
 Entrez Trn
 Entrez Vmbr
 Entrez Wt
 Entrez Ync
 Entrez Ync2
 Entrez Ync3
 Entrez Ync4
 Entrez Ync5
 Entrez Ync6
 Entrez Ync7
 Entrez Ync8
 Entrez Ync9
 Entrez Ync10
 Entrez Ync11
 Entrez Ync12
 Entrez Ync13
 Entrez Ync14
 Entrez Ync15
 Entrez Ync16
 Entrez Ync17
 Entrez Ync18
 Entrez Ync19
 Entrez Ync20
 Entrez Ync21
 Entrez Ync22
 Entrez Ync23
 Entrez Ync24
 Entrez Ync25
 Entrez Ync26
 Entrez Ync27
 Entrez Ync28
 Entrez Ync29
 Entrez Ync30
 Entrez Ync31
 Entrez Ync32
 Entrez Ync33
 Entrez Ync34
 Entrez Ync35
 Entrez Ync36
 Entrez Ync37
 Entrez Ync38
 Entrez Ync39
 Entrez Ync40
 Entrez Ync41
 Entrez Ync42
 Entrez Ync43
 Entrez Ync44
 Entrez Ync45
 Entrez Ync46
 Entrez Ync47
 Entrez Ync48
 Entrez Ync49
 Entrez Ync50
 Entrez Ync51
 Entrez Ync52
 Entrez Ync53
 Entrez Ync54
 Entrez Ync55
 Entrez Ync56
 Entrez Ync57
 Entrez Ync58
 Entrez Ync59
 Entrez Ync60
 Entrez Ync61
 Entrez Ync62
 Entrez Ync63
 Entrez Ync64
 Entrez Ync65
 Entrez Ync66
 Entrez Ync67
 Entrez Ync68
 Entrez Ync69
 Entrez Ync70
 Entrez Ync71
 Entrez Ync72
 Entrez Ync73
 Entrez Ync74
 Entrez Ync75
 Entrez Ync76
 Entrez Ync77
 Entrez Ync78
 Entrez Ync79
 Entrez Ync80
 Entrez Ync81
 Entrez Ync82
 Entrez Ync83
 Entrez Ync84
 Entrez Ync85
 Entrez Ync86
 Entrez Ync87
 Entrez Ync88
 Entrez Ync89
 Entrez Ync90
 Entrez Ync91
 Entrez Ync92
 Entrez Ync93
 Entrez Ync94
 Entrez Ync95
 Entrez Ync96
 Entrez Ync97
 Entrez Ync98
 Entrez Ync99
 Entrez Ync100
 Entrez Ync101
 Entrez Ync102
 Entrez Ync103
 Entrez Ync104
 Entrez Ync105
 Entrez Ync106
 Entrez Ync107
 Entrez Ync108
 Entrez Ync109
 Entrez Ync110
 Entrez Ync111
 Entrez Ync112
 Entrez Ync113
 Entrez Ync114
 Entrez Ync115
 Entrez Ync116
 Entrez Ync117
 Entrez Ync118
 Entrez Ync119
 Entrez Ync120
 Entrez Ync121
 Entrez Ync122
 Entrez Ync123
 Entrez Ync124
 Entrez Ync125
 Entrez Ync126
 Entrez Ync127
 Entrez Ync128
 Entrez Ync129
 Entrez Ync130
 Entrez Ync131
 Entrez Ync132
 Entrez Ync133
 Entrez Ync134
 Entrez Ync135
 Entrez Ync136
 Entrez Ync137
 Entrez Ync138
 Entrez Ync139
 Entrez Ync140
 Entrez Ync141
 Entrez Ync142
 Entrez Ync143
 Entrez Ync144
 Entrez Ync145
 Entrez Ync146
 Entrez Ync147
 Entrez Ync148
 Entrez Ync149
 Entrez Ync150
 Entrez Ync151
 Entrez Ync152
 Entrez Ync153
 Entrez Ync154
 Entrez Ync155
 Entrez Ync156
 Entrez Ync157
 Entrez Ync158
 Entrez Ync159
 Entrez Ync160
 Entrez Ync161
 Entrez Ync162
 Entrez Ync163
 Entrez Ync164
 Entrez Ync165
 Entrez Ync166
 Entrez Ync167
 Entrez Ync168
 Entrez Ync169
 Entrez Ync170
 Entrez Ync171
 Entrez Ync172
 Entrez Ync173
 Entrez Ync174
 Entrez Ync175
 Entrez Ync176
 Entrez Ync177
 Entrez Ync178
 Entrez Ync179
 Entrez Ync180
 Entrez Ync181
 Entrez Ync182
 Entrez Ync183
 Entrez Ync184
 Entrez Ync185
 Entrez Ync186
 Entrez Ync187
 Entrez Ync188
 Entrez Ync189
 Entrez Ync190
 Entrez Ync191
 Entrez Ync192
 Entrez Ync193
 Entrez Ync194
 Entrez Ync195
 Entrez Ync196
 Entrez Ync197
 Entrez Ync198
 Entrez Ync199
 Entrez Ync200

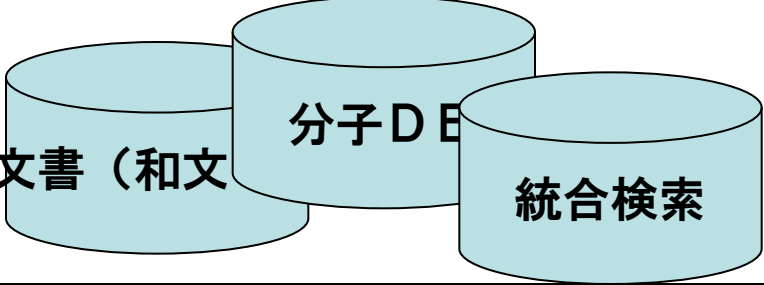
肝細胞癌
 胆管細胞癌
 胎児性肝癌
 CCl4肝癌
 etc



(例) 名称にp53を含む遺伝子ヒットだけで42個)



ステップ2. 明瞭な専門概念用語による一般的な検索 索引のマージ



統一形式 (開発中)

- <生物名辞書:Taxonomy>
学術用語集の取り込み(終了)
NCBI-Taxonomi 日本語対訳中
- <解剖辞書:Ontology>
Foundation Model of Anatomy (ontology)
Terminologia Anatomica船戸訳(辞書)
- <疾患名>
ICD-10, Medical textbooks, OMIM

ポータル整備運用：H19版リリース後の目標と進行状況

6月2日にDBCLSで開発した23のサービスを公開

サービス/目標	進行状況
横断検索/ DB追加(目標200)と検索の向上	<ul style="list-style-type: none"> ・生命科学系DB横断検索については順調にDBを追加しており、KEGG DRUG, PIRなど重要なDBの他、微生物DBの検索を8月以降オープンした。 ・特許公報のみであった特許検索に公開・公表特許公報を追加し公開する(バイオJapanにあわせて) ・タンパク関連DBの横断検索実現のために、索引を作成中。 ・検索精度の向上について、現状各レコードへのランク付けは難しいので、利用者が望みのDBについて検索結果を取捨選択できるようにする。 ・ユーザインタフェースの改良→絞込みがわかりづらい、何由来のDBかわかりにくいなど ・索引のアップデートなど管理インタフェースを作製して今後の維持管理を容易にする。
生命科学系DBカタログ/ DB追加(目標150)と充実	<ul style="list-style-type: none"> ・生命科学系DBカタログのDB追加は、NARのDB issue からデータを作製中である。 ・既存レコードの記載内容に関する追加、修正は今のところできていない。
生命科学学会・協会検索	<ul style="list-style-type: none"> ・年会情報(日時、ウェブサイトなど)の情報を調査して追加 ・各学会の発行する学会誌のオープンアクセス状況を調査し利用者に提示、アクセスを容易にするための機能を追加する。
日本語文献 検索/ 文献の追加、他	<ul style="list-style-type: none"> ・生物物理学会に年会要旨と論文誌の利用の許諾をいただいたので12月の生物物理学会をめどに追加する。 ・それ以外は今後、努力が必要であり、上記学会・協会検索と連動してオープンアクセス誌の利用を考えたい。 ・クリエイティブコモンズやサイエンスコモンズなどの活動の普及と連携していく
全般	<ul style="list-style-type: none"> ・今年度は各コンテンツ内のデータの追加、充実化を図ることが重要 ・デザイン、使用感の向上、統一感が必要 ・当初、枠組みのインタフェースをgadgetの作製やAPIとの連携が容易なようにデザインしたが、サービス毎の開発状況からまだ有効に利用できる段階では無い ・メニューがわかりにくいとの指摘も多く改良必要 ・検索やツールのサービスが多いのでSEOを対策する

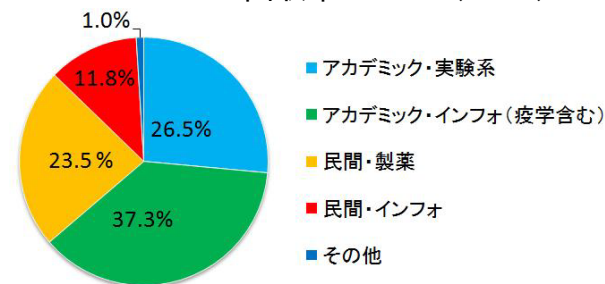
19年度成果に対するユーザー評価まとめ

●実施概要

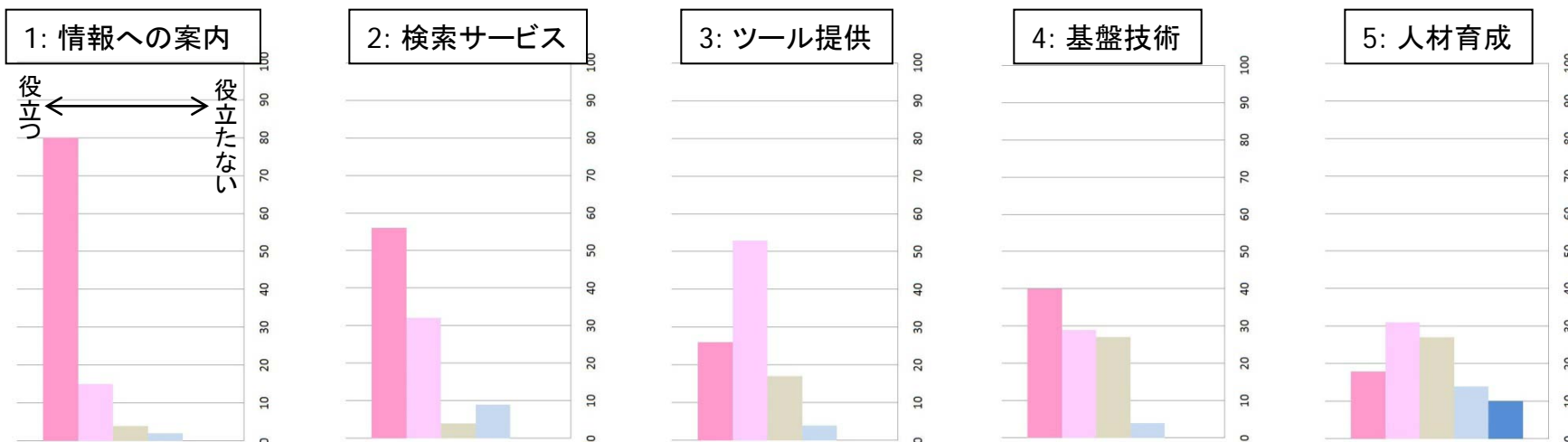
- ・H19年度成果に対するウェブサイト記入方式のユーザ評価を7/22～8/5に実施
- ・評価記入サイトには、5つの項目(代表的な17のサービスを分類)について選択肢とコメント欄を用意
- ・同サイトには、サービスの紹介と内容の理解のために、テストクエリや画面サンプル等を用意

●回収率および評価者の背景

回答者数: 92名 (2名辞退)
回収率: 62% (57/92)



●目的ごとに分類したサービス内容についての全般的な評価(5段階)



●フリーコメントの記述内容

- ・各分類全般についてあるいは個別のサービスに対して(191件、一部重複あり)
- ・サービス全般に対するフリーコメントや今後当プロジェクトで提供すべきサービスについて (28件)

統合データベース講習会AJACS

<http://motdb.DBCLS.jp/>

1. データベースを使いこなすための技術的な基礎知識から効率的な利用法について広く知ってもらう
 - 特に統合データベースプロジェクトによって作成・維持管理されているサービスを使いこなすための効率的な利用法について広く知ってもらうことを目的
2. データベースを作成、維持管理していく仕事に関しても興味を持ってもらう
3. 参加者全員がコンピュータを使いながらの講習
 - 講習会ウェブサイト(MotDB: <http://motdb.dbcls.jp/>)の活用→復習が容易
 - これまでに102コンテンツ作成してきた統合TVも活用しながらのユニークな講習形態



2008年度開催の統合データベース講習会(3回)

日時、呼称	場所	参加者
2008/05/30 AJACS湘南	日本大学 生物資源科学部	64人
2008/07/03 AJACS本郷1	東京大学工学部 12号館(DBCLS)	20人
2008/08/29 AJACS蝦夷	北海道大学 情報科学研究科	30人



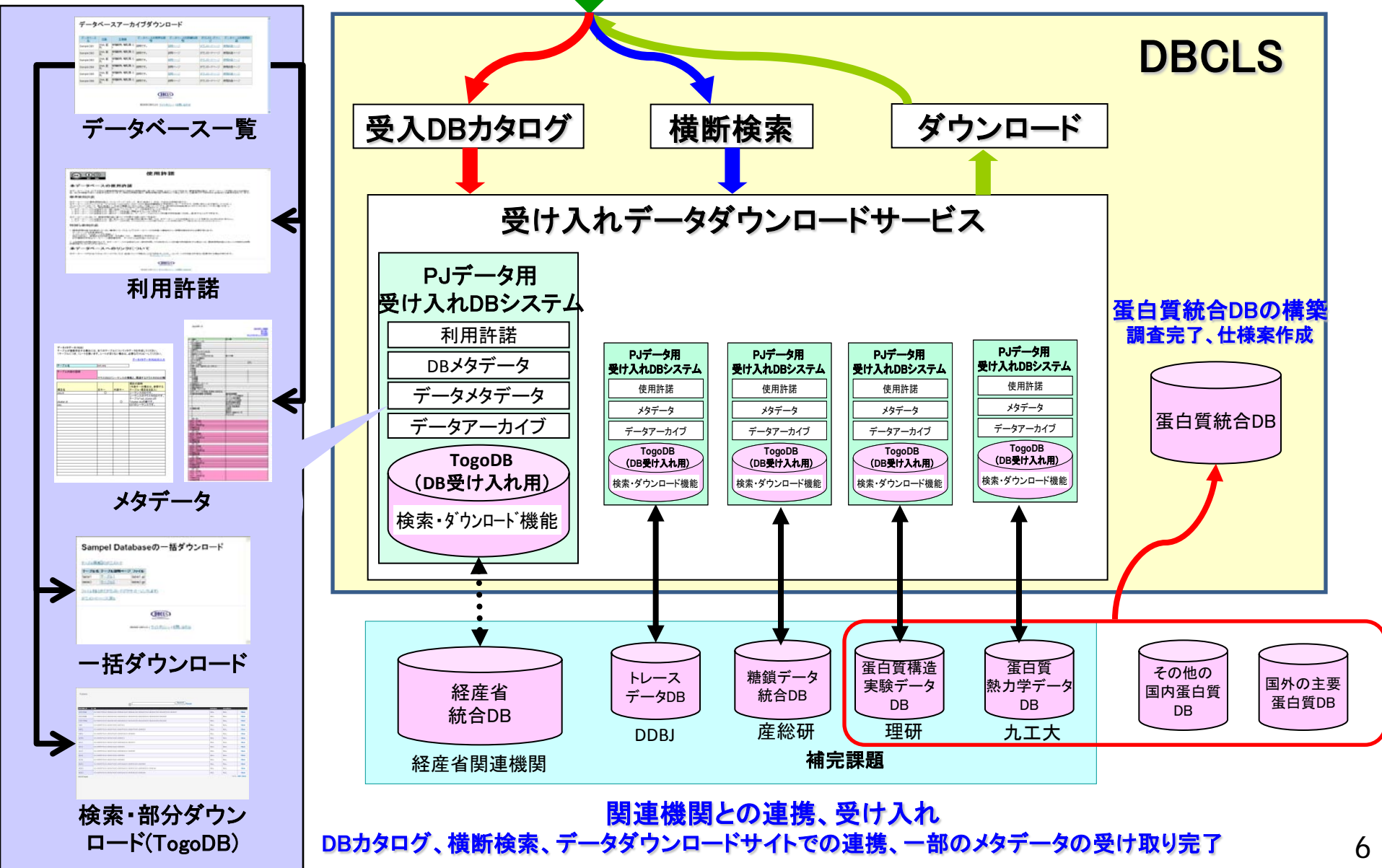
AJACS 蝦夷の様子



あと4回開催予定(九大、長浜バイオ大、東工大、DBCLS)

データベースの受入と運用 進行状況

受け入れデータダウンロードサービス
基盤構築
第1バージョン完成



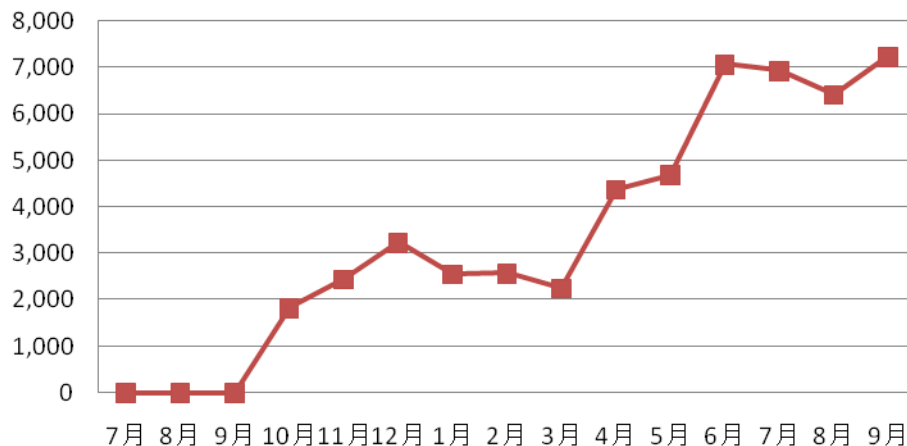
関連機関との連携、受け入れ
DBカタログ、横断検索、データダウンロードサイトでの連携、一部のメタデータの受け取り完了

ポータルサービスのアクセス統計

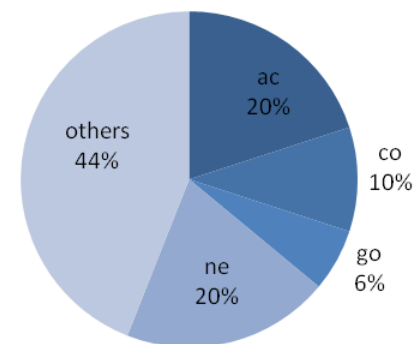
解析概要

- 公開から本年9月の間のアクセス統計のうち訪問者数とアクセスされたページ数を示した。
- ログの解析にはAWstatsを利用した。

ユニーク訪問者数



ページビューのドメイン内訳



- othersにはDNS逆引き不可のものを含む

今後の予定

参加機関にアクセスログ収集を依頼

(目的) 本プロジェクトの有用性の評価

(方法)

1. Apache NCSA combined log formatで収集
2. 生データあるいは解析データをDBCLSへ送付
(個人情報保護の関する問題点検討後、決定)

ページ数

