

引用可能 URL プロジェクト

石黒真木夫@統計数理研究所

1. 電子情報は「印刷論文」にはるかに勝る有用性を秘めている
2. にもかかわらず、対応する「雑誌論文」を欠く電子情報の形でのみ公表された情報は「論文」と同格の評価を受けない。
3. その大きな理由のひとつは、現状では電子情報が安定性を欠き、「参考文献」として引用することが出来ない点にあると考えられる。
4. 「引用」が可能であるためには
 - (a) 常に「そこ」にあること。
 - (b) 預けた本人でも書き換えできないこと。が必要かつ充分である。

1. 電子情報を安心して「引用」出来る事になり、学界に対する貢献は非常に大きいと考えられる。
2. 電子情報の引用が社会的に認知されるようになれば、現在の学術雑誌、学術図書のありかたを根本から変える契機となるのは明らかであり、
3. 将来的には情報社会の様相を革新することが期待される。

安定した運営組織が必須である。
運営組織の役割

1. サーバ管理
 - (a) 遠隔バックアップ
 - (b) セキュリティ管理
 - (c) ハードウェア更新
2. ユーザー/ファイル管理
 - (a) 収録ファイルの選定 = 「引用可能 URL」資源の配分
 - (b) 100年後に読めるであろうファイルタイプの見極め
3. Public Relations
 - (a) 「引用可能 URL」の重要性に関する広報活動
4. ブラウザ管理
 - (a) 保有ファイルを読むためのブラウザの提供？
 - (b) 現行ブラウザで読めるように保有ファイルを変換？

プロジェクト進行状況

1 プロジェクト進行状況

1. 2002年度統計数理研究所プロジェクト研究「統計科学情報の高度利用」の一部として統計数理研究所内で認知され、実施経費の割り当てをうける。
2. 「統計科学情報の高度利用」研究会 (2003.2.3-4)
3. 宿久洋氏 (鹿児島大学) と「遠隔バックアップシステム」の実験をすることを合意。
4. 「分散型論文管理システム」を発注
5. pdf 版 AISM の提供を依頼

Subject: pdf 版 アナルスデータセットに関してのお願い

Date: Sun, 02 Mar 2003 01:13:01 +0900

北川 AISM 編集委員会委員長殿：

平成 15 年度のプロジェクト経費（「統計情報の高度利用に関する研究」（馬場さん））をいただいて「引用可能 URL プロジェクト」を開始致しました。

第一期分の作業として「分散型論文管理システム」の構築が無事に進行中で、近々下記の要領で試験運用を行えるようになると思います。

つきましてはそこに pdf 版 アナルスデータセットを置かせて頂きたく、許可とご支援をお願いします。実際に支援をお願いするのは雑誌編集室の渡邊さんになるかと思いますが、お願いの具体的な内容は

1. ネットワーク、あるいは適当なメディアを介してのデータセットの提供、
2. データフォーマットに関する情報の提供

の 2 点になると考えています。ご検討下さい。

引用可能 URL プロジェクトに関してはプロジェクトのホームページをご覧ください。石黒のホームページ（<http://www.ism.ac.jp/~ishiguro>）の「projects」というリンクをクリックすると「PROJECTS」というページに飛んでそこに「URL」というリンクがあります。これをクリックするとユーザー名とパスワードが要求されます。

ユーザー名:polar

パスワード:star

を入力して下さい。

石黒真木夫

PS:polar star = 常に「そこ」にあるもの、です。

PPS: この件は

>>>> On Fri, 20 Dec 2002 17:27:00 +0900, "Genshiro Kitagawa" <kitagawa@ism.ac.jp> said:

>>> 下に添付したメールのように、AISM のうち統計数理研究所で出版した部分、

>>> Vol.1-39 および Supplement について編集室で Web 公開の準備が完了いたしました。

>>> ました。

>>> つきましては、統計数理研究所の Web で公開することについての許可を頂

>>> きたく、広報委員会にてご審議のほどよろしくお願い申し上げます。

と似ていますが，別のことなので混同されないようご注意ください．

記

利用の目的： 「分散型論文管理システム」の構成および運用に関して実際的な問題点を洗い出す事．

利用者の範囲： 当面馬場プロジェクト参加者のうちから当プロジェクトに関心を持つ研究者，石黒 + 馬場が責任を持って選んだメンバーの範囲で，試験的に利用する．
利用者の名簿は馬場プロジェクト間で公表する．

試験期間： 当面 2004 年度末までとするが，試験の結果を見て，試験期間の延長，利用者の範囲の拡大などの形で実用運転に漸進させていく予定．

結果の公表： 馬場プロジェクトの結果の公表の一部として公表するが，
適当な機会をとらえて本プロジェクトについての発表も行う．

以上

「分散型論文管理システム」システム要求仕様/作業手順 案

I. システム要求仕様

1. 最小限 2ヶ所に分散したサブシステムからなるシステムとする．将来，サブシステムの数が増やせる設計．
2. 各サブシステムはそれぞれスタンドアロンで，WEB サーバ/データベースサーバとして機能する事を要する．
3. データベース機能は，全文検索，キーワード指定での検索，期間を限定した検索が可能であること．（期間に関しては，和暦，西暦のいずれも作れること）しぼり込み検索，AND/OR の論理が作れる事．

フィールド構成は

雑誌名
タイトル
著者名
巻数
発行年
号数
始ページ 終ページ
(pdf) ファイル名
ファイルサイズ
フリーメモ領域

4. 可能なら，ユーザーはどちらのサブシステムにアクセスしているのか意識する必要がないように構築する．

5. サブシステムの一部がダウンした場合には、もう一方に残った情報にもとづいて全システムを再構築出来ることが必要である。
6. 各サブシステムのローカルなバックアップについてもある方が望ましい。
7. security 管理に十分に注意を払う。
サブシステムのそれぞれが異なる OS のもとで動いているというのでもいいかもしれない。
8. コンテンツは
 - A 論文又はそれに準ずる「著作」であって、電子化されているもの。
たとえば PDF(テキストデータ+フォントの形で全文検索可能であるものを原則とする)
 - B 会員名簿、原則として URL のみ「会員」の引越し先を追跡するシステム。
(<http://www5.ocn.ne.jp/~qmss/>>参考サイト!!)
9. 各「論文」コンテンツに URL が与えられること。
10. コンテンツのローディングはバッチ一括と個別ローディングが可能である事。
11. 個別ローディングへのキューは公開で誰でも入力可。
ただし、審査員の判定を経て公開される仕組みを備えている事。
12. バッチ処理で不備が発見されたコンテンツは個別ローディングのキューに入る事。
13. 最新のセキュリティー情報に応じた迅速な処置が可能である事。

(まだ何か抜けているかもしれません)

II. 作業手順

年度内に 2.1 まで実施

1. メインサーバーの立ち上げ
 - 1.1 データベースの設計, インストール/実験
 - 1.2 個別コンテンツ入力処理 インターフェース
インストール/実験
 - 1.3 バッチ入力処理インターフェース
インストール/実験
(AISM の入力を予定. 傷のあるデータの入力もテスト)
2. (ホロン社内において) 分散サーバーの立ち上げ, 実験.
 - 2.1 外部からのアクセスの実験
 - 2.2 ユーザーによる個別コンテンツ入力処理の実験
 - 2.3 メインサーバーダウンから回復のテスト
 - 2.4 サブサーバーダウンから回復のテスト
 - 2.5 プロジェクト名による「恒久的 URL」の取得 (作業手順としてはもっと前?)
3. 鹿児島大学に分散サーバーを設置/テスト

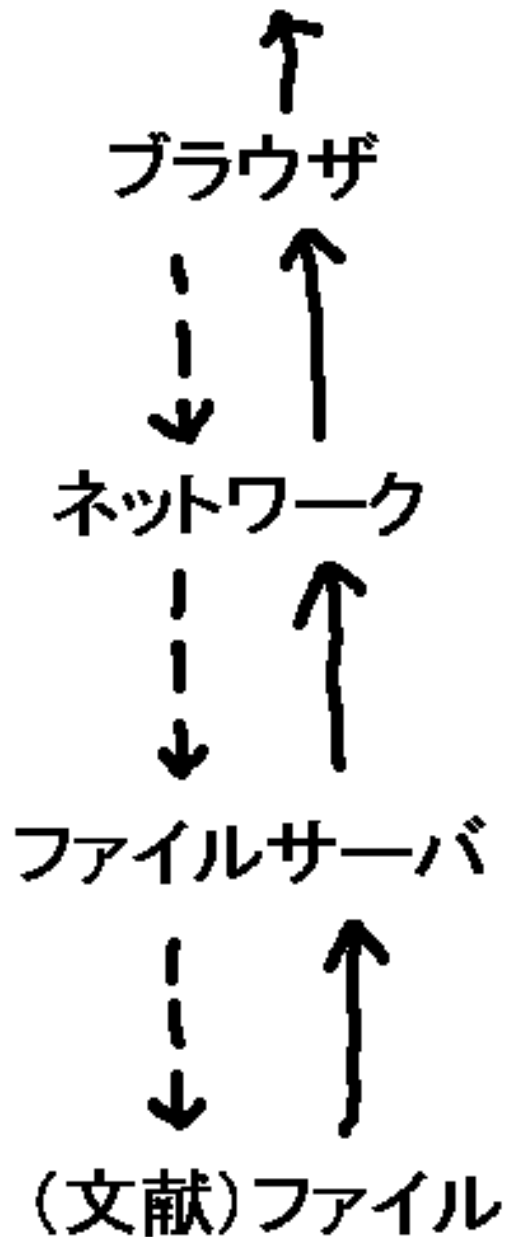
引用可能 URL プロジェクト

石黒真木夫@統計数理研究所

2002年度統計数理研究所プロジェクト研究による研究会「統計科学情報の高度利用」

日時 2003年2月3日(月)～4日(火)

場所 統計数理研究所講堂



引用可能 URL
49.5-914.2-414.2-919-119-319-619-9
運営組織 (案)

4.5-54.5-99.5-29.5-

まとめ

プログラム

<<<第1日 2月3日(月)>>>

【セッション1 10:00 - 12:00】

座長 田中豊(岡山大学\\

)

1. 馬場康維(統計数理研究所)

統計情報の高度利用の研究について

2. 栗原考次(岡山大学)

空間情報の構造分析とHotspots 検出

3. Ji-Hyun Kim (Soongsil University, 統計数理研究所客員)

An extension of the log-rank test

【セッション2 13:00 - 15:00】

座長 道家暎幸(東海大\\

学)

4. 佐藤美佳(筑波大学)

ファジィクラスタリング評価とその応用

5. 清水信夫(統計数理研究所)

WWW上の統計関連情報とその表示

6. 石黒真木夫(統計数理研究所)

引用可能 URL プロジェクト

【セッション3 15:20 - 17:20】

座長 後藤昌司(大阪大\\

学)

7. 宿久洋(鹿児島大学)、橋口博樹(目白大学)

類似度に基づく大量データの表示

8. 中川重和・尾高好政・小林和真(倉敷芸術科学大学)

HTTPリクエストデータの解析

9. 山本義郎(多摩大学)・森裕一(岡山理科大学)・宿久洋(鹿児島大学)

データ・解析事例公開のためのデータ記述とモデルの記述

【セッション4 総合討論 17:30 - 18:00】

座長 馬場康維(統計数理研究所\\

)

討論者 後藤昌司(大阪大学)、田中豊(岡山大学)、垂水共之(岡山大学)\\

、氏家勝巳(東海大学)、林篤裕(大学入試センター)

<<<第2日 2月4日(火)>>>

【セッション5 10:00 - 12:00】

座長 垂水共之(岡山大\\

学)

10. 金藤浩司(統計数理研究所)

環境科学と統計科学の融合への第一歩

11. 石橋雄一(スタットラボ)

統計チュートリアル現在の現在と今後の発展

12. 柳久男(岡山理科大学)・垂水共之・飯塚誠也・笛田薫・藤野友和(岡山大学)

Web上の統計教育のための教材

連絡先 栗原考次(岡山大学)、馬場康維(統計数理研究所)

1. 電子情報は「印刷論文」にはるかに勝る有用性を秘めている
2. にもかかわらず、対応する「雑誌論文」を欠く電子情報の形でのみ公表された情報は「論文」と同格の評価を受けない。
3. その大きな理由のひとつは、現状では電子情報が安定性を欠き、「参考文献」として引用することが出来ない点にあると考えられる。
4. 「引用」が可能であるためには
 - (a) 常に「そこ」にあること。
 - (b) 預けた本人でも書き換えできないこと。が必要かつ充分である。

「引用可能な電子情報システム」が成功すれば

1. 電子情報を安心して「引用」出来る事になり、学界に対する貢献は非常に大きいと考えられる。
2. 電子情報の引用が社会的に認知されるようになれば、現在の学術雑誌、学術図書のあるかたを根本から変える契機となるのは明らかであり、
3. 将来的には情報社会の様相を革新することが期待される。

運営組織(案)

1. サーバ管理
 - (a) 遠隔バックアップ
 - (b) セキュリティ管理
 - (c) ハードウェア更新
2. ユーザー/ファイル管理
 - (a) 収録ファイルの選定 = 「引用可能 URL」資源の配分
 - (b) 100年後に読めるであろうファイルタイプの見極め
3. Public Relations
 - (a) 「引用可能 URL」の重要性に関する広報活動
4. ブラウザ管理
 - (a) 保有ファイルを読むためのブラウザの提供？
 - (b) 現行ブラウザで読めるように保有ファイルを変換？

ブラウザ管理

1. 保有ファイルを読むためのブラウザの提供？
2. 現行ブラウザで読めるように保有ファイルを変換？

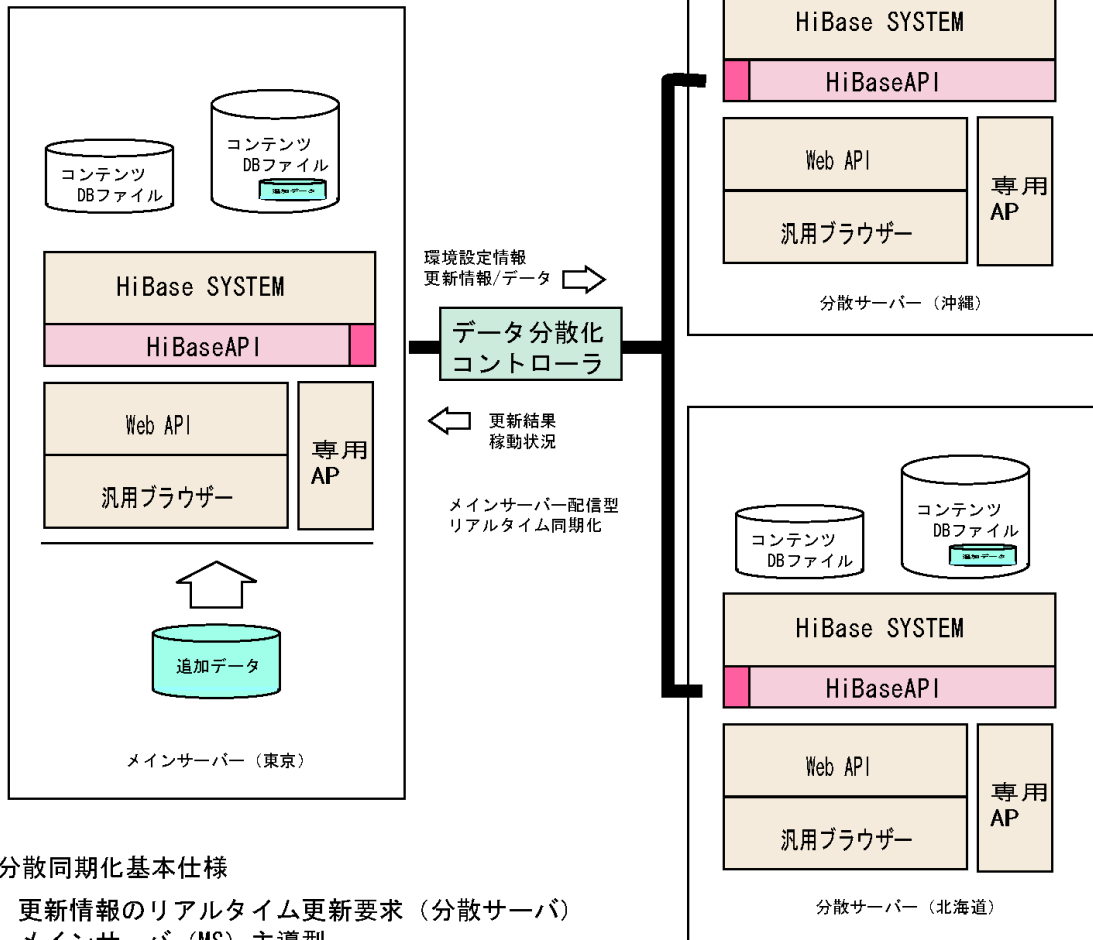
Public Relations

1. 「引用可能 URL」の重要性に関する広報活動

組織的サーバ管理

1. 遠隔バックアップ
2. セキュリティ管理
3. ハードウェア更新

HiBaseコンテンツ分散管理手法



分散同期化基本仕様

- 更新情報のリアルタイム更新要求（分散サーバ）
- メインサーバ（MS）主導型
- MSの切替可
- 起動時の同期確認（更新ログレベル）
- 同期ログの作成（MSでは分散サーバログも）
- MSは分散同期の定義の変更時に配信（同期）
- 定義の変更権限の設定
- 稼動状況の監視（起動時/更新要求時）
- メインサーバの交代（定義の変更/配信をリアルタイムに実施）
- 同期障害のアラーム（更新同期が正常でない場合にどの分散サーバかをアラームする）

組織的ユーザー / ファイル管理

1. 収録ファイルの選定 = 「引用可能 URL」資源の配分
2. 100年後に読めるであろうファイルタイプの見極め

Web 上での数式表現の利用例

$$P(y|x) = \frac{P(x,y)}{P(x)}$$

x が与えられたときの y の条件付分布は同時分布を周辺分布で割ることで求められる。