

リポジトリの統計

慶應義塾大学 五十嵐健一

2012.12.14

XooNlps研究会2012 福井大学WS

リポジトリの統計とは？

◎ どんなもの？（用途）

- なぜ必要？どんなふうに使われる？具体的には？
 - 基本的には一定期間蓄積して、特定の基準で集計して、比較することになる
 - 機関リポジトリ自体を知るための統計と利用を比較するための統計が良く使われる

◎ どんなもの？（手段、具体的にどうやって？）

- どこで取っている？取りまとめている？
 - 統計ごとに異なる。比較や集計の基準も異なる。
- 生ログは？
 - 基本的にはhttpサーバのログが元となる
- 分析ツールは？Webalizer、Google Analytics、ROAT
 - どの統計に、どの数字を使うのか、事前に検討する
 - 一度基準を決めたら、同じ基準で比較する（基準を変えた場合、その変化の前後で比較はできなくなる）

リポジトリの統計とは？

- ◎ 目的：
機関リポジトリ自体、またそこで扱っているコンテンツを
いろいろな角度から解析するための指針・指標として利用
する・される
 - 日本国内に機関リポジトリはいくつある？
 - この機関リポジトリに登録されているコンテンツの内訳は
どうなっている？
 - この機関リポジトリは、どのくらい・どう使われている？
 - リポジトリ内のコンテンツはどこからどうやって利用される？
- ◎ 手段：
 - 目的を達成するために、どんな手段(ツール)を利用するか
 - コストはどのくらいかかるか？
 - どういった基準・集計の機関を適用するか
- ◎ できれば運用初期から必要となる統計を意識して、目的と
手段の組み合わせを設計しておくが良い

機関リポジトリを知るための統計とは？

- リポジトリの紹介やアクセス誘導のため
(NIIのCSI委託事業サイト、JAIRO、
Open DOARなど)
 - リポジトリの集合リストにデータを登録し、それをリポジトリ自体の紹介とする。また、そこからのリンクによりアクセスを引き込む。
 - リポジトリの属する集合リストの中で同じ基準で比較することで、そのリポジトリの特徴である、コンテンツ登録種別の割合、登録数の伸び率などを把握・紹介するために利用出来る。

機関リポジトリの利用を知るための統計とは？

◎ 利用されていることの証明

- 利用統計とも、アクセスログとも。アクセス数、ダウンロード数、などと言い換えられることも。
 - ただ、基準がないと比較が難しく、意味がない統計になってしまう。
- 機関リポジトリの統計に関する基準。ROAT、PIRUS (COUNTER) など。
- ROATは千葉大学のCSI委託事業のプロジェクト。
- PIRUSはJISCというイギリスの機関リポジトリ関連の仕様などを検討しているグループが規定している。COUNTERを機関リポジトリに適用するための仕様
 - COUNTERとは、電子ジャーナルの利用され方を同じ基準で比較できるようにしたもの。

利用例として

- 世界の機関リポジトリは、Open DOARで確認可能

<http://www.opendoar.org/>

- 日本の機関リポジトリに関するの取りまとめは、NIIでリスト化されている

<http://www.nii.ac.jp/irp/archive/statistic/>

- 日本のリポジトリ・コンテンツはJAIROに詳しい

<http://irdb.nii.ac.jp/analysis/index.php>

- ただ、自機関のリポジトリの直接の統計は？
→ 自機関でやるしかない！？

自機関のリポジトリの統計は 何故必要？どう使われる？

◎ 各種調査

- 文部科学省 学術情報基盤実態調査

http://www.mext.go.jp/b_menu/toukei/chousa01/jouhoukiban/1266792.htm

- 朝日新聞出版 大学ランキング

◎ 運用方針の決定

- ハードウェアの選定に利用
- コンテンツの利用状況確認に使用

リポジトリの統計情報集計ツール？

◎ Webalizer、AWStats

- オープンソースの汎用的なツールを使う
 - 利点→情報源が豊富。無償で試せる。
 - 欠点→標準のまま使う？カスタマイズする？どうやって？
→明確な比較基準が定まっていない。

◎ ROAT

- 機関リポジトリアウトプット評価システム
(Repository Output Assessment Tool) も提供してきました。AWStatsというwebサーバーのアクセスログ分析システムを、ガイドラインに沿ってカスタマイズしています。

◎ Google Analytic

- Googleが提供するアクセス解析サービス

◎ XooNips標準の統計機能

デモ

- ◉ Webalizer
- ◉ AWStats
- ◉ ROAT
- ◉ Google Analytics
- ◉ KOARAの統計ツール(XooNIps標準)

まとめ

- 機関リポジトリの利用統計は、機関リポジトリが実際にどう使われているか？理解するための指標となるので、直接のユーザー・サービスではないが、重要である
- 目的(何のための指標とする?)と手段(ツールは?基準は?期間は?)を明確化する
- コストパフォーマンスを考える。
 - 何を目的として、それは本当に必要なのかどうか。
 - すでに管理団体があるものは、登録するだけでも良い。しかし、詳細は自機関でやる必要も。
 - どんな統計でも、集計作業には、それなりのコスト・労力が必要
 - できるだけ早期に、できれば運用前から考えておけると良い