

P2Pアーカイブ

~P2P型Webアーカイブシステムの構築~

札幌 寛之

森下 睦

InternetArchive

- 「インターネットの図書館」
 - 全てのインターネットコンテンツを永久に保存
 - 過去の状態も参照可能
 - しかも無料



未踏/セキュリティ・キャンプ

未踏事業ポータルページ

▽ [未踏事業とは](#) | ▽ [公募情報](#) | ▽ [イベント・TOPICS情報](#) | ▽ [未踏の世界](#) |

未踏事業とは

「未踏事業」は、ITを駆使してイノベーションを創出することのできる独創的なアイデアと技術を有するとともに、これらを活用する優れた能力を持つ、突出した若い人材を発掘・育成することを目的としています。

2000年度から「未踏ソフトウェア創造事業」として開始し、2008年度より若い人材の発掘・育成に重点化すべく再編した「未踏IT人材発掘・育成事業」として現在実施しています。2つの事業を総称して「未踏事業」と呼んでいます。

「未踏事業」で輩出してきた延べ1,600人もの優れたクリエイターたちは、既に様々な分野で活躍しています。



未踏/セキュリティ・キャンプ

> [未踏事業](#)

> [セキュリティ・キャンプ事業](#)

2015年度公募
未踏

詳しくは...

InternetArchiveの問題

データ量

- 2017年現在: **26**ペタバイト
 - データ量は**今後**も増え続ける
- これらを提供する**通信帯域**も必要
- データ量の問題で**動画ファイル**は現在未保存

寄付による運営

- 現在、運営費を全額寄付に依存
 - 今後機能し続けられる保証が無い
- 今後、カナダに新たなアーカイブを構築予定
 - 500万ドル (5.5億円) かかる見込み
 - もちろん、これも寄付

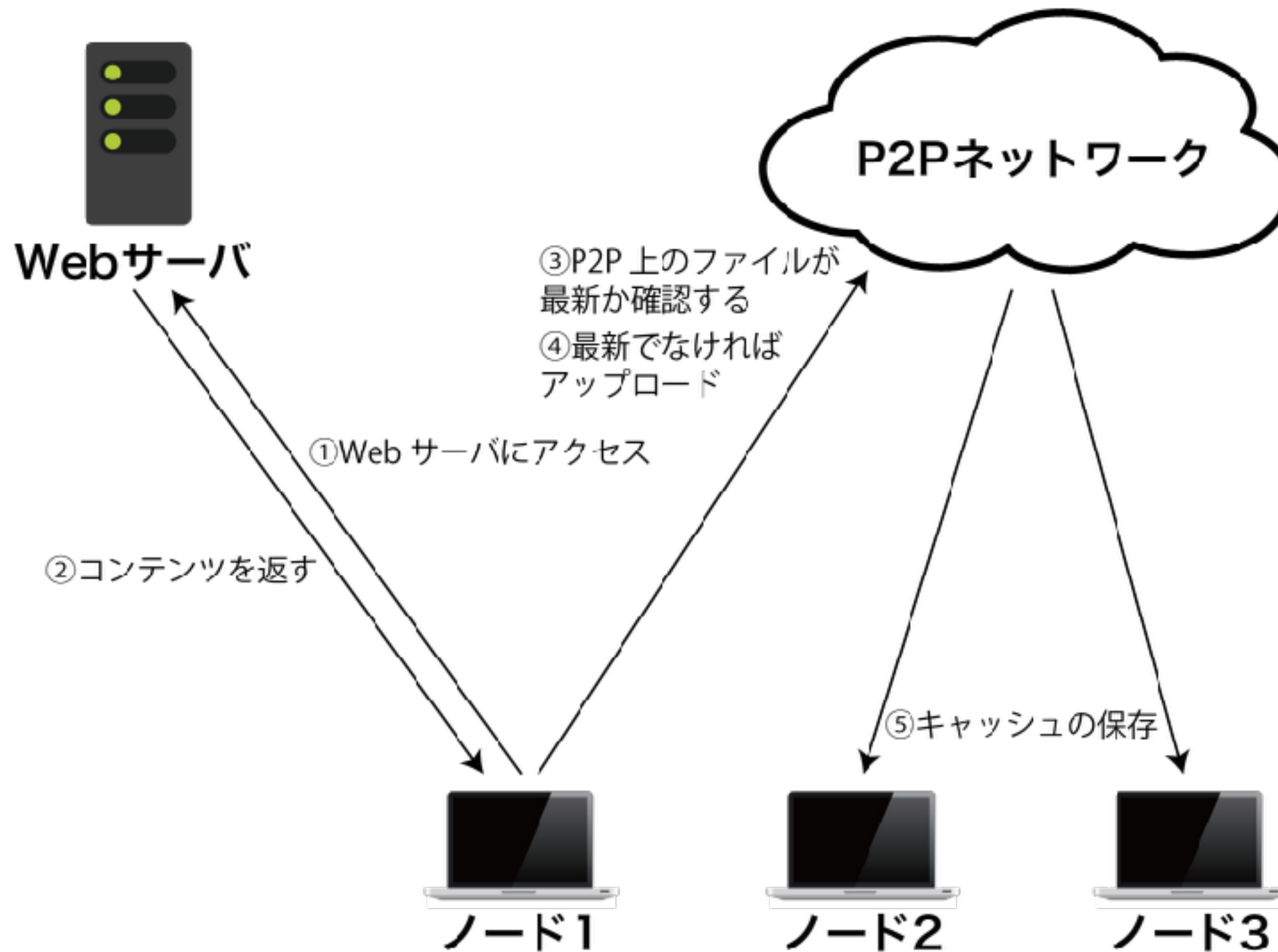
どうすればいい？

P2P

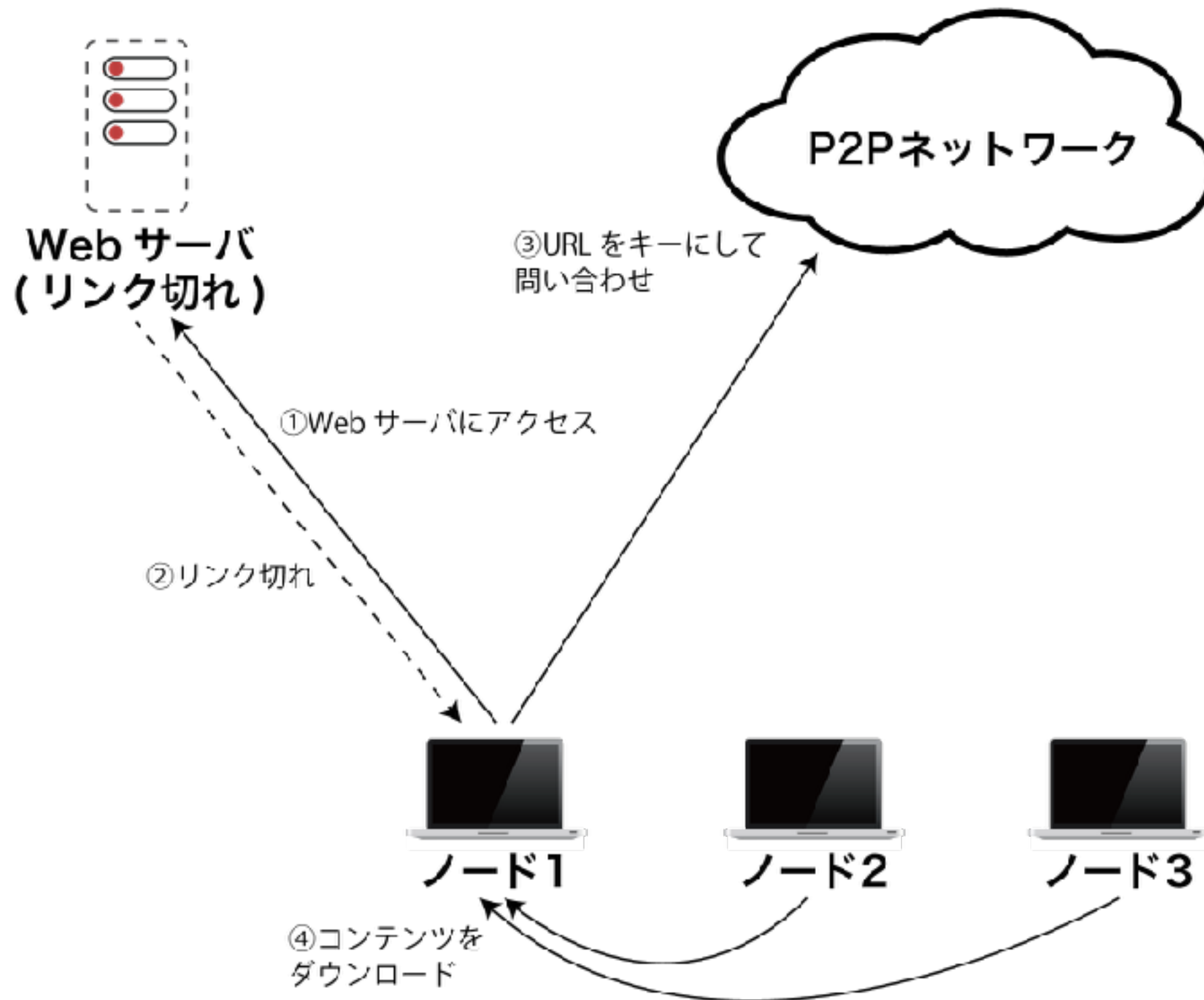
提案内容

- P2Pネットワーク上でInternetArchiveを実現
 - 各ユーザ = クローラ & ホスト
 - ユーザに負荷分散

P2Pネットワークへのアップロード



P2Pネットワークからのダウンロード



提案法により解決する問題

- データ量の問題

→ 各ノードに負荷を分散することで解決

- 運営費の問題

→ 運営母体が強力なサーバを持つ必要が
なくなり解決

CDNとしての機能

- P2Pネットワークからデータをダウンロード
→ 既存のサーバに負荷をかけずに
高速にダウンロードできる可能性



類似ソフトウェア

WikiP2P

- WikipediaをP2Pネットワーク上でホスティング
- 相違点
 - 使用可能なWebサイトをWikipediaに限定
 - 過去のバージョンは見られない
 - CDNとしての利用は想定していない

CacheP2P

- P2PネットワークでCDNを実現
- 相違点
 - サイト運営者が設定したサイトのみ使用可能
 - 現在サイトに接続しているユーザとのみ通信可能
 - 過去のバージョンは見られない

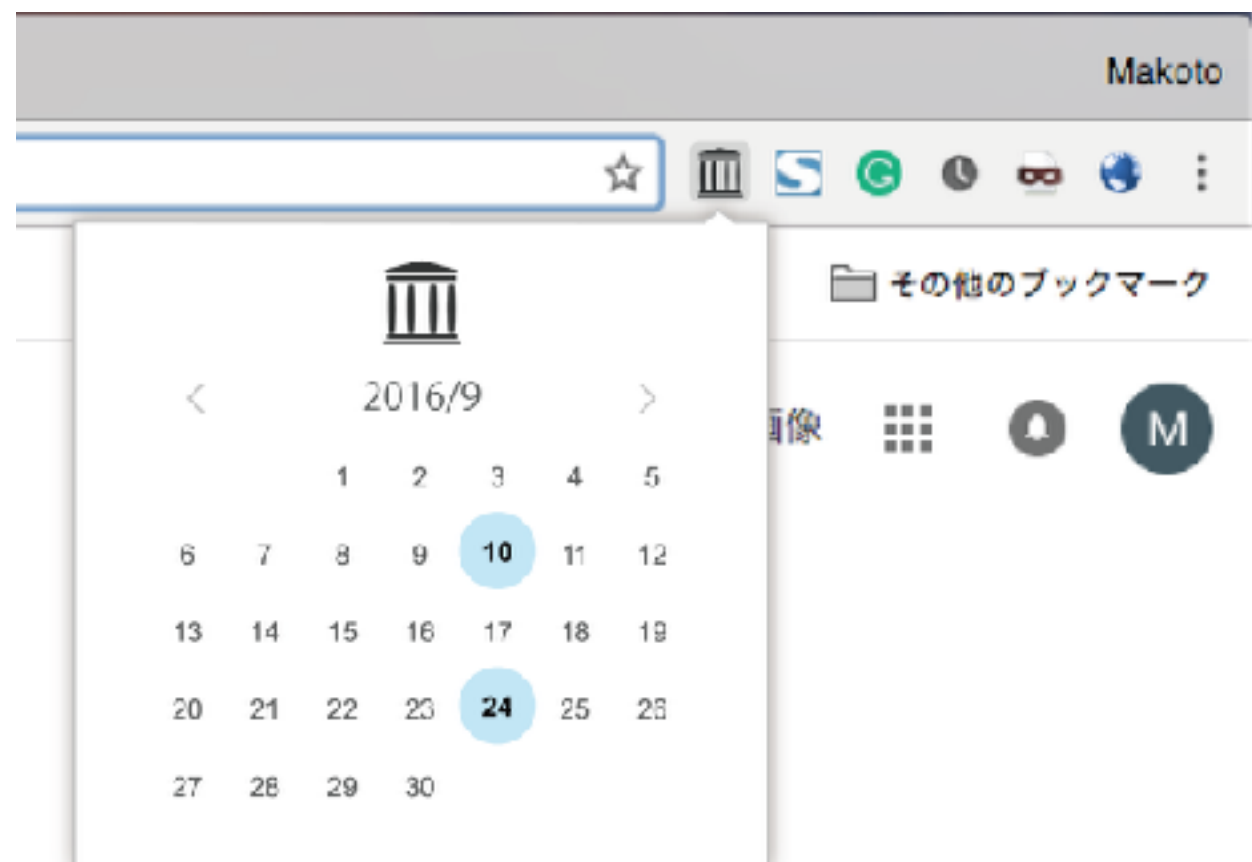
Freenet

- P2Pネットワーク上で
Webサイトをホスティング
- 相違点
 - 既存のインターネットとは分離
 - 匿名性などに重きを置き、転送効率は重視せず
 - 過去のバージョンは見られない

未踏プロジェクトで目指す
ゴール

ブラウザプラグインとして動作

- 簡単にインストールでき、
多くのユーザの参加が期待できる
- わかりやすいUI



著作権の問題

- 日本では著作権を無視してデータをアップロードできない可能性
 - 未踏プロジェクトでのゴールは許可されたWebサイトのみを対象に動作

管理できるP2Pネットワーク

- 純粋なP2Pネットワークは管理者が存在せず、データの削除が難しい
→ Winny等の問題が発生
- 本プロジェクトは後から管理者によってデータの削除が可能なネットワークを構築

未踏後のゴール

- 多くのユーザを獲得
 - ほぼ全ての情報がP2Pネットワーク上に流れる
 - Webの構造がP2Pネットワークに移行する

使われ方， 打ち出し方

- フリーソフト， オープンソースでの提供
- 本プロジェクトで得られたノウハウを
ブログ等の形で提供
- 将来， 国会図書館での事業にも活用可能
 - 国会図書館もWebArchiveの構築を試行中

役割分担

- 札幌寛之
 - ブラウザプラグイン
 - P2Pネットワークの構築
- 森下睦
 - P2Pネットワークの構築
 - 論文サーベイ

現在の進捗

- 管理可能なP2Pネットワークを構築する手法をサーベイ中
- P2Pネットワーク上へのウェブサイトのアップロード、ダウンロードを実装
 - ただし、まだブラウザプラグインでは無い

まとめ

- 従来のWebArchiveをP2P上で実現
 - 負荷分散, 可用性向上等のメリット
 - CDNとしても利用可能
- ブラウザプラグインとして動作
- 管理できるP2Pネットワークを構築
 - 後からファイルの削除が可能

END