

2006 年度未踏ソフトウェア創造事業
コンテンツ検索エンジン ZDEA 2.0 の開発

開発代表者：山岡 幸作

共同開発者：山澤 宏鑑

担当 PM：河野 恭之

プロジェクト管理組織：テクノロジーシードインキュベーション(株)

1 背景及び目的

近年、クロック周波数 1GHz を超える高速 CPU、大容量ハードディスクドライブやメモリ、光ファイバー回線の普及によって、創作コストが大きく低下し、ワールド・ワイド・ウェブでは、マルチメディア作品に接する機会が増加した。

しかし、Flickr や YouTube のような一カ所にアップロードされた写真や動画を公開 API 経由での検索は可能ではあったが、ウェブに散在するマルチメディア作品は検索することは未だ困難である。

なぜならば、HTML 文書に記述されたメタデータとマルチメディア作品のバイナリー・ファイルの関連づけが完全ではないからである。

そこで、Google 画像検索のように作品単位で検索するのではなく、クリエイターが公開しているウェブサイトという単位で検索することを発案した。

本プロジェクトの目的は、クリエイターのウェブサイト検索が従来よりも容易になることで、クリエイターのウェブサイトへの閲覧者を増やし、クリエイターの創作活動を活性化させることで、さらなる作品の創造を促すことである。

2 開発の内容

ウェブ・アプリケーションとして、一般的な LAPP によるオープンソース・ソフトウェアを用い、本プロジェクトのシステムを構築した。

OS	CentOS(Red Hat Enterprise Linux クローン)
ウェブサーバ	Apache
RDBMS	PostgreSQL
プログラム言語	PHP

本プロジェクトによるシステムは、以下の 2 台のサーバで構成されている。



検索結果画面

検索サーバ

クローラ

The screenshot shows the ZDEA search results page for the keyword 'ハルヒ'. The page displays 122 results. The first result is 'でんでんはうす' (Denden House) with 1680 items and a 5-star rating. It includes a thumbnail of the site and a list of keywords: (343) 花 (271) メカ (261) 海 (259) 風 (258) ガンダム (221) 少女 (191) 男 (188) 水着 (184) キュラクター (183) 猫 (178) 雪 (158) 電 (157) メガネ (155) 赤 (149) by (149) 車 (149) 学生 (148) 写真 (147) 愛 (146) 画像 (143) ロボ (142) 魔法 (138) 桜 (127) アニメ (125) og (125) 美 (121) 朝 (120) ギャラリー (120) はね (120) 恐竜 (120) ゲーム (119) おこ (118) 女の子 (117) 太陽 (110) 山 (110) http (107) 物語 (107) 羽 (106). Below this, the second result is 'こもれびのーと' (Komorebi no Oto) with 910 items and a 5-star rating. It also includes a thumbnail and a list of keywords: 月 (765) メイド (185) 絵 (149) 人 (113) 漫画 (68) 和 (58) 電 (45) コミティア (44) 更新 (41) 犬 (39) ゲーム (39) アニメ (38) 同人 (31) 香 (30) 東方 (30) 花 (28) 風 (28) 猫 (25) 耳 (24) 秋 (24) 空 (24) 男 (19) エロ (19) スクール水着 (18) 学生 (16) キュラクター (16) 魔理沙 (16) このみ (16) ネコミミ (15) 女 (15) Fate (15) パンツ (14) toheart2 (14) og (14) イルファ (14) 萌え (14) 東 (13) 作 (13) 少女 (13) ああ (13) ガンダム (12) イラスト (12) 姉 (11) key (11) 少年 (11) 春 (11) 18禁 (11) 魔法 (10) 朝 (10) 女の子 (10) 音. The page also has a search bar, navigation links (about, blog, help?), and pagination (1 2 3 4 5 6 7 8 9 10 →).

サイト名
公開イラスト枚数
五段階評価
サイト画面

著作権
引用条件に則した
サムネイル表示
自動抽出した
キーワード(本文)

3 従来の技術(または機能)との相違

既存のサービスと比較しながら、本プロジェクト成果物の特徴を簡潔に述べる。

- ・クリエイターのウェブサイトの特化している。
- ・評価基準の客観性が高く、公平である。
- ・検索利用者にとって、表示情報が多く、検索しやすい。
- ・既存サービスと比較して、クリエイターの負担が少ない。

3.1 Google との比較

Google と比較して、クリエイターのウェブサイトの特化しており、クリエイターのウェブサイトを検索することについては、本プロジェクト成果物のほうが優位にある。

クリエイターのウェブサイトの多くは frame タグを多用していること、ディープリンクを禁止していることなど、検索エンジンに最適化されたウェブサイトと比較して、Google に評価されにくい構造のため、クリエイターのウェブサイトは一般のウェブサイトと埋もれやすく、検索しにくい。

3.2 Yahoo!ディレクトリとの比較

専門家(Yahoo!社員)によって、登録されたウェブサイトを検索できるが、非公開の基準で登録され、客観性に欠けている。本プロジェクト成果物によるウェブサイトの評価は客観性が高く、クリエイターが納得できるだけの公平さがあり、クリエイターの信頼を得ることが可能である。

また、検索結果の一覧表示は、五十音順、もしくは人気順(基準不明)であり、ウェブサイトのタイトル名だけではどのようなウェブサイトか伺い知ることはいできない。

本プロジェクト成果物は、ウェブサイトの規模による客観的な数値と、スクリーンショット(画面画像)、サムネイル(縮小画像)、メタデータを備えており、検索利用者は容易に検索が可能である。

3.3 既存のクリエイター検索サービスとの比較

本プロジェクト成果物は、既存のクリエイター検索サービス(例 tinami.com 等)と比較して、クリエイターの負担が少ない。(クリエイターが最適なメタデータを付与しようと思わなければ、特に負担は発生しない)

既存のクリエイター検索サービスは、クリエイター自身がIDとパスワードを申請し、手でメタデータを入力し、更新のたびにメタデータを更新する必要がある。また、数ヶ月に一回しかメタデータを更新できないという制約が存在する。

また、検索結果は、更新日時順であるため、ウェブサイトの作品の品質と、検索結果の順位の相関性が低く、ウェブサイトへ移動して後悔することが多かったが、本プロジェクト成果物は、前述の通り検索利用者にとって、とても使いやすいものになっている。

4 期待される効果

波及的に活性化される産業分野はコンテンツ産業であり、個人の求める作品を、個人が制作する活動をサポートしたいと考えている。

5 普及(または活用)の見通し

9 月に公開後、毎日 1500 人(ロボットを除いたセッション数)の検索利用者が訪れ、合計 5000 回(ロボットを除いたページビュー)利用されている。今後も開発を継続することで、さらに利用を促したい。

6 (参照)

開発者 URL <http://www.zdea.com>

以上