

「地域面検索β版」(全文検索可能)に切り抜きイメージを追加

記事データベース「朝日新聞クロスサーチ」の「地域面検索β版」(1991年から1996年まで全文検索可能、2024年2月リリース)について11月25日、以下の2つの新機能を追加いたしました。

- ◆ 「切り抜きイメージ」を表示
- ◆ 紙面に記事の位置を示す「赤いカギマーク」を表示

●「地域面検索β版」の画面

The screenshot shows the search results page for '松阪の豪商屋敷 復元へ' (Restoration of Matsukage's Merchant House). The search criteria are set to '全国の地域面' (National Regional Pages). The results list three items, with the first one selected. The detailed view shows the article text and a '切り抜きイメージ' (cutout image) of the article's content. A red key mark is visible on the page image, indicating the article's position.

●「地域面検索β版」(紙面デジタル復刻)

- ・スキャンした新聞紙面をAIが解析。自動で記事や見出しの領域を検出し、それぞれを並び替えて記事全文をOCRで読み取りテキスト化しました。

- ・記事の分類(92項目)もAIが自動で付与。

- ・「全国の地域面」(1935年～1999年)のうち、1991年3月から1996年12月までの約6年分の地域面約31万ページから、記事・広告390万件のキーワード検索が可能になりました。

●「地域面検索β版」の扱いについて(2024年11月25日以降)



「地域面検索β版」のご利用はオプションB「全国の地域面」のご契約が必要になります。

01

記事領域取得

画像処理・機械学習により、紙面画像から記事構成要素の座標・分類を取得

02

OCR(光学文字認識)

文字情報を含む領域にOCRを適用
ノイズ処理、画像の加工により精度UP

03

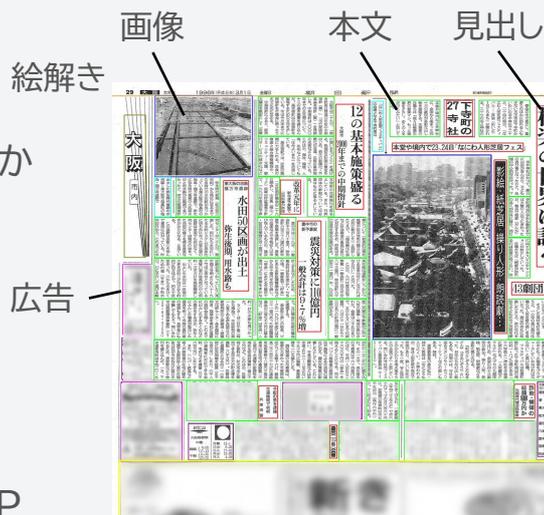
記事連結

新聞ルール・画像処理・自然言語処理により、見出し・本文・画像・絵解きをグルーピング & 本文の読み順特定
位置情報取得

04

メタデータ付与

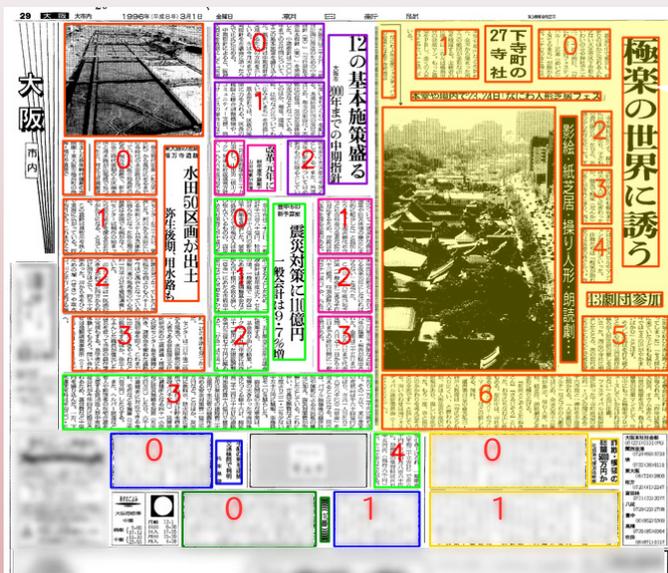
自然言語処理により、記事の分類推定や著作権・個人情報の判定を実施



<領域取得>



<本文の連結>



見出し:極楽の世界に誘う 本堂や境内で23、24日「なにわ人形芝居フェス」...

本文:都会の中で孤立し、墓参りのほかに訪れる人も少ない天王寺区下寺町かいはいの寺や神社で二十三、二十四の両日、「なにわ人形芝居フェスティバル冥途(めいど)INてらまち」が開かれる。...

分類:演劇

出典)朝日新聞, 1996年3月1日