

# 全文検索による試験問題検索 システム

---

—新規作成問題の類似文書検索を中心として—

石岡恒憲, 橋本貴充, 大津起夫  
(大学入試センター)

# 概要

---

- 背景
- システム要件
- 検索の実際
- まとめ

# 試験問題の点検は困難

---

- 教科・科目数が多い
- H24年度より地歴・公民で時間割の区別なし
  - 「世界史A」「世界史B」「日本史A」「日本史B」「地理A」「地理B」「現代社会」「倫理」「政治経済」「倫理・政治経済」の10科目から最大2科目
- H27年度から理科
  - 理科①「物理基礎」「化学基礎」「生物基礎」「地学基礎」4科目
  - 理科②「物理」「化学」「生物」「地学」4科目
  - 理科②移行措置「理科総合A」「理科総合B」「物理Ⅰ」「化学Ⅰ」「生物Ⅰ」「地学Ⅰ」の6科目

# コンピュータ支援は不可欠

---

- 試験問題の重複や類似性を検討
  - 自らの担当教科だけでなく
  - 過去の試験分
  - 比較的近年の有名国公立大学の入試問題

# 既存試験問題の検索システム

---

- キーワード検索 (Google検索のようなもの)
  - 洩れなく、網羅的に
  - 縦書き、PDF
- 新作問題 (Word, Excel, PDF) の類似検索
  - 全体としての類似性 < 洩れのなさ
  - キーワードの自動抽出 & 即時の検索
- 共通
  - 違和感のない応答

# システム要件

---

- Hyper Estraier(古いフランス語で「迷う」)
  - 定評のある全文検索システム
  - N-gramを拡張したインデックス方式→洩れの無さ
  - コード変換ライブラリ,データ圧縮ライブラリ,組み込み用DB
  - Linux2.2以降
- 形態素解析Mecab
- 若干のカスタマイズ、ノウハウ
  - PDFの索引付け

# キーワード検索

http://ae01.rd.dnc.ac.jp/cgi-t Search Result: 小選挙区...

ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(I) ヘルプ(H)

小選挙区比例代表並立制 Search advanced help

10 per page, with , order by , clip by --

Results of 1 - 10 of about 27 for 小選挙区比例代表並立制 (0.253 sec.)\*

**現代社会**

**2011-C1-T-02-z-03.pdf** (10066)  
第1問 次の文章を読み、下の問い(問!~#)に答えよ。(配点 14) 現代の!民主主義においては、選挙が民...べて  
小党を排除しやすくなるとされる。#日本の衆議院の**小選挙区比例代表並立制**の議席は、小選挙区180議席・比  
例代表300議席という配分...  
<http://ae01.rd.dnc.ac.jp/DNC/2011T/C1/2011-C1-T-02-z-03.pdf> - [detail] - [similar]

**倫理、政治・経済**

**2013-C4-T-04-z-02.pdf** (9835)  
第1問 次の文章A,Bを読み、下の問い(問!~#)に答えよ。(配点 14) A戦後の日本の政治は、憲法第1条と日...改  
正された。この改 正によって衆議院議員の選挙制度が**小選挙区比例代表並立制**となった結果、政党内部で執  
行部の力が強まり、2001年に成...  
<http://ae01.rd.dnc.ac.jp/DNC/2013T/C4/2013-C4-T-04-z-02.pdf> - [detail] - [similar]

**2007-C1-H-01-z-04.pdf** (9075)  
第1問 次の文章を読み、下の問い(問!~")に答えよ。(配点 22) 日本における!民主政治と法の支配の関...次  
の!~\$のうちから一つ選べ。!!衆議院においては、**小選挙区比例代表並立制**が採用されており、衆議院議員の  
定数のうち、小選挙区選出...  
<http://ae01.rd.dnc.ac.jp/DNC/2007H/C1/2007-C1-H-01-z-04.pdf> - [detail] - [similar]

# キーワード検索

ブラウザのアドレスバー: <http://ae01.rd.dnc.ac.jp/cgi-t> Search Result: 小林秀雄

メニュー: ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(I) ヘルプ(H)

検索欄: 小林秀雄 Search advanced help

10 per page, with , order by , clip by --

Results of 1 - 10 of about 23 for 小林秀雄 (0.205 sec.)\*

**国語**

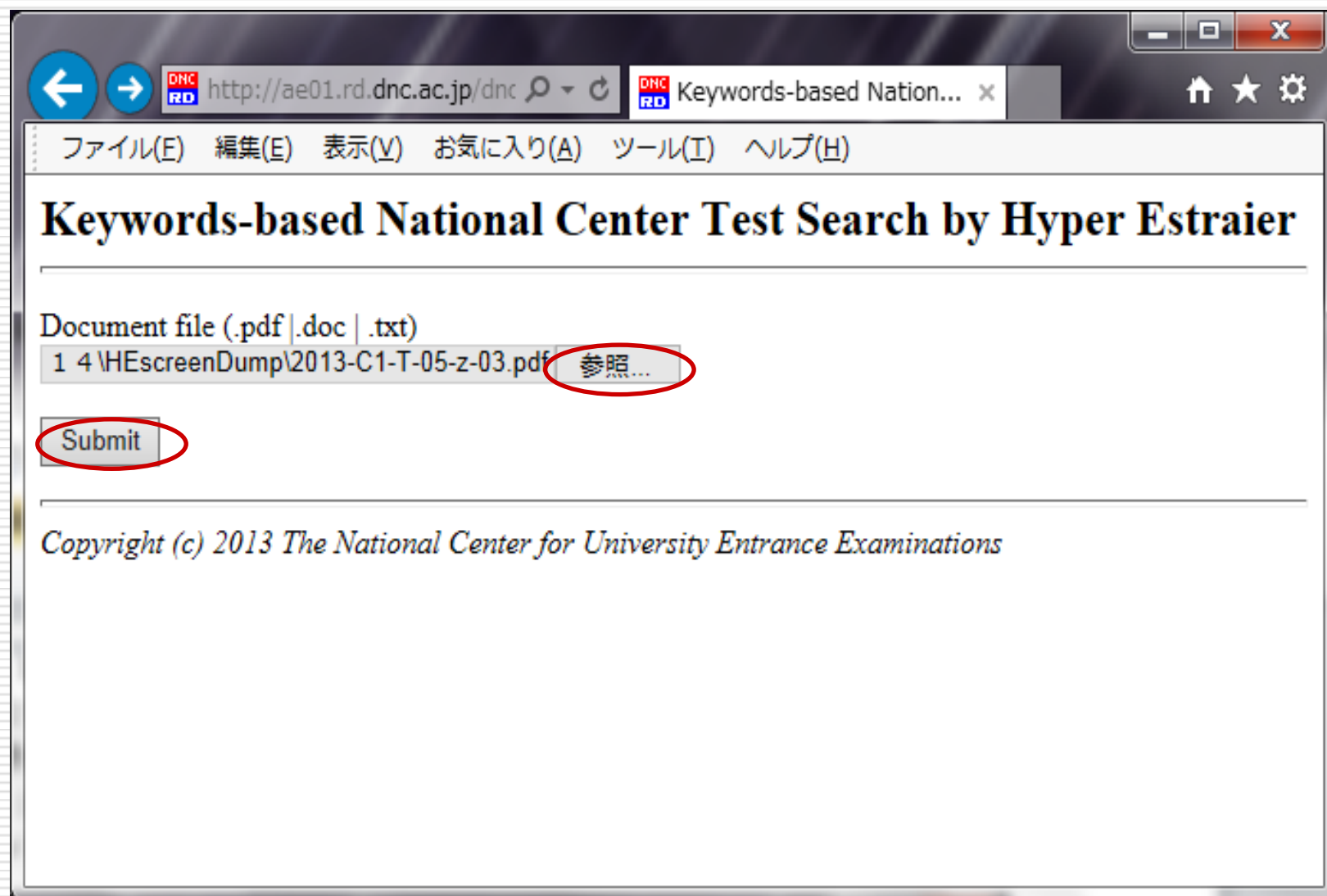
[2006-A1-T-01-z-05.pdf](#) (8886)  
第1問 次の文章を読んで、後の問い(問!~#)に答えよ。(配点 50) 美術館 ... 説を取り上げている。たとえば「注# ~ 小林秀雄 こつとうだ。」美の近代的鑑賞「を ... 四八 ~」。\$モデルニテ ... 近代性。%小林秀雄 ... 批評家「一九〇二 ~ 一九八三 ~ ... 董いぢり」という鑑賞態度を称揚した小林秀雄と異なり、ポール・ヴァレリーはそ ...  
<http://ae01.rd.dnc.ac.jp/DNC/2006T/A1/2006-A1-T-01-z-05.pdf> - [detail] - [similar]

**倫理**

[2007-C2-T-03-z-05.pdf](#) (8698)  
第1問 次の文章を読み、下の問い(問!~#)に答えよ。(配点 24) 答えが見つからなくてもどこまでも自己 ... あるということを、仁斎は改めて発見した。近代の批評家小林秀雄は、初期の評論において「批評の対象が己れであると他人 ...  
<http://ae01.rd.dnc.ac.jp/DNC/2007T/C2/2007-C2-T-03-z-05.pdf> - [detail] - [similar]

[2007-C2-T-03-z-02.pdf](#) (8510)  
第1問 次の文章を読み、下の問い(問!~#)に答えよ。(配点 24) 答えが見つからなくてもどこまでも自己 ... あるということを、仁斎は改めて発見した。近代の批評家小林秀雄は、初期の評論において「批評の対象が己れであると他人 ...

# 関連文書検索の初期画面



# 関連文書検索

The screenshot shows a web browser window with the address bar displaying `http://ae01.rd.dnc.ac.jp/cgi/` and a tab titled "Keywords-based Nation...". The browser's menu bar includes "ファイル(E)", "編集(E)", "表示(V)", "お気に入り(A)", "ツール(I)", and "ヘルプ(H)".

The main content area is titled "Keywords-based National Center Test Retrieval Output". It displays two search results for the keyword "ねじれ国会" (60 hits).

**Keyword (HIT): ねじれ国会 (60)**

**URI:** <file:///var/www/html/DNC/2011H/C3/2011-C3-H-02-z-05.pdf>  
第1問 次の文章を読み、下の問い(問!~#)に答えよ。(配点 19) 日本の国会は!二院制を採用している。... 能させるための課題としてしばしば指摘されてきた。%「ねじれ国会」が生じると参議院の存在感は高まるが、それでもなお、参... 問 5 下線部!に関連して、2007年から2009年まで続いた「ねじれ国会」についての記述として最も適切なものを、次の!~\$のうち...

**URI:** <file:///var/www/html/DNC/2011H/C3/2011-C3-H-02.pdf>  
第1問 次の文章を読み、下の問い(問!~#)に答えよ。(配点 19) 日本の国会は!二院制を採用している。... 能させるための課題としてしばしば指摘されてきた。%「ねじれ国会」が生じると参議院の存在感は高まるが、それでもなお、参... 問 5 下線部!に関連して、2007年から2009年まで続いた「ねじれ国会」についての記述として最も適切なものを、次の!~\$のうち...

**Keyword (HIT): アメリカ (5544)**

**URI:** <file:///var/www/html/DNC/1986T/TG/1986-TG-T-03.pdf>  
第3問 アングロアメリカとラテンアメリカに関する下の問い(問 1~5)に答えよ。(配点 20)  
問 1 次... の図の線①~⑤のうちから、アングロアメリカとラテンアメリカの境界として最も適切なものを一つ選べ... 。20 問 2 アングロアメリカとラテンアメリカの気候について述べた次の文①~⑤のうち... から正しいものを一つ選べ。21 ① 両アメリカでは、冷帯気候に属する

# システムの保守管理

---

- インデックスの更新
- 削除文書の反映
  - インデックスの削除
- 更新作業の自動化
  - 定期的に実行(cron)
  - 現在夜間→1時間おきも容易

# おわりに

---

- Hyper Estraierはコマンドベースで起動可能
  - 自作スクリプト
  - UIFを含むシステム改良が容易
- 検索エンジンの中身がわかる安心感
- NIIからWiki見出し語辞書の提供
- 改良へのコメント