

統語・意味解析コーパス (NPCMJ) チュートリアル

セッション3

検索インターフェースの基本操作

(NPCMJ Development Interfaces)

アラステア・バトラー、長崎郁

2022.3.4

NPCMJ Development Interfaces

- NPCMJ を検索するためのインターフェース
 - 文字列 (KWIC) 検索
 - Tgrep2/Tregex の検索式を使ったツリー検索
- 現在、487 files、98,312 trees、1,455,020 words
 - 不定期に解析ミスを修正

1

セッション3の内容

1. ファイル一覧
2. KWIC 検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果のダウンロード
6. インターフェースに関するドキュメント

2

セッション3の内容

1. ファイル一覧
2. KWIC 検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果のダウンロード
6. インターフェースに関するドキュメント

3

ファイル一覧：入口

NPCMJ Development Interfaces を開く



Corpora hosted by the Oriental Institute, University of Oxford, UK

This page gives access to a growing number of related parallel corpora that share the same annotation principles as well as a wider access to grammatical functions. Some of the corpora offer support for an automatic creation of logical expressions with discourse relations, age, including quantitative classes, logical connectors, and product relations.

The data analysis of this page is in a continual process of being hand worked. We strive to reduce error rates as much as possible. We have no been but continued developments is not only desirable for the correction and engagement in the data that it brings, but it is also perfectly along with having actively used on-line resources. The interface access real-time "living" data, is completely up-to-date working files, and links a relative to frequent updates.

There is no completion target for the annotation. There is rather an appetite for ever improving data analysis, and the insights that this can offer a multitude of exciting purposes.

Old Japanese

- The Oxford/NINJA Corpus of Old Japanese (ONCOJ) (Project Front page)

The ONCOJ has been ongoing since 2011 (until 2017 under the name "The Oxford Corpus of Old Japanese (OCOJ)"). It was in conceived of and designed as a resource for linguistic research, but it also includes features useful for work in history, literature, culture, & in general, the ONCOJ contains the full corpus of Old Japanese poetic texts, including the Manyōshū.

MC307's (also MC307) (Project Front page) - this annotation for the first stanza seven poems of the Manyōshū

This corpus fragment from 2016 was an early pilot study into applying rich parsing principles to the poems of the Manyōshū, a complete text containing some of the oldest attested forms of the Japanese language.

Contemporary Japanese

- The Kansai Textbank (Project Front page) - the under-development main corpus of the NINJA Parallel Corpus of Modern Japanese (NINJA-PC)

The Kansai Textbank has been ongoing since 2011 (until 2016, the corpus was known under the name "The Kansai Textbank"). The name changed when its development became the main focus for the NPCMJ project at the National Institute for Japanese Language and Linguistics (NINJAL). The initial a goal is to provide the best of both worlds: a resource that best facilitates pattern search for linguistic research. Singlehit version - corpus have been released on the NINJA Parallel Corpus of Modern Japanese (NINJA-PC) on a yearly basis since 2016.

- The Kansai Textbank (Project Front page)

The Kansai Textbank is created by exchanging the Japanese script morphological base of the Kansai Textbank for a different (transmission) morphological base while retaining the syntactic level annotation. Currently just under a third of the data of The Kansai Text is used.

ファイル一覧：Overview (概要) ページ

Toggle overview tree | source | ttp | newspan | indexing | flora | graph | download [Front page external link] Help

Tree search interface

Files:

TGrep2 search expression:

Search basic Toggle tag set TGrep2 search guide

487 files; 98,312 trees; 1,455,020 words ← ファイル数、ツリー数、語数

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4191	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9327	essay : Chiri, Mashihō (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1762	essay : Chiri, Mashihō (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	966	essay : Chiri, Mashihō (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashihō (1960). "Aikokushin" <i>Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3035	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3
8 aozora_Dazai-2-1940	447	5037	fiction : Dazai, Osamu (1940). <i>Hashire Merousu</i> . Shincho, May edition
9 aozora_Doyle-1905	784	10533	fiction : Doyle, Arthur Conan - Kareha (trans.) (2000). <i>Noouddo no ken</i>

ファイル一覧：Overview (概要) ページ

Toggle overview tree | source | ttp | newspan | indexing | flora | graph | download [Front page external link] Help

Tree search interface

Files:

TGrep2 search expression:

Search basic Toggle tag set TGrep2 search guide

487 files; 98,312 trees; 1,455,020 words

それぞれのファイルの：
 ファイル番号
 ファイル名 (Filename)
 ツリー数 (tree count)
 語数 (word count)
 内容 (content description)

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4191	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9327	essay : Chiri, Mashihō (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1762	essay : Chiri, Mashihō (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	966	essay : Chiri, Mashihō (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashihō (1960). "Aikokushin" <i>Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3035	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3
8 aozora_Dazai-2-1940	447	5037	fiction : Dazai, Osamu (1940). <i>Hashire Merousu</i> . Shincho, May edition
9 aozora_Doyle-1905	784	10533	fiction : Doyle, Arthur Conan - Kareha (trans.) (2000). <i>Noouddo no ken</i>

ファイル一覧：Overview (概要) ページ

Toggle overview tree | source | ttp | newspan | indexing | flora | graph | download [Front page external link] Help

Tree search interface

Files:

TGrep2 search expression:

Search basic Toggle tag set TGrep2 search guide

487 files; 98,312 trees; 1,455,020 words

それぞれのファイルの：
 ファイル番号
 ファイル名 (Filename)
 ツリー数 (tree count)
 語数 (word count)
 内容 (content description)

ファイル番号をクリック
 → ファイルの中身を表示

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4191	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9327	essay : Chiri, Mashihō (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1762	essay : Chiri, Mashihō (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	966	essay : Chiri, Mashihō (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashihō (1960). "Aikokushin" <i>Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3035	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3
8 aozora_Dazai-2-1940	447	5037	fiction : Dazai, Osamu (1940). <i>Hashire Merousu</i> . Shincho, May edition
9 aozora_Doyle-1905	784	10533	fiction : Doyle, Arthur Conan - Kareha (trans.) (2000). <i>Noouddo no ken</i>

Context view (コンテキスト表示) ページ

Toggle context view Go to Kusunoki Corpus overview

title	Torokko
author	Akutagawa, Ryunosuke
date	1922
source	Aozora Bunko (https://www.aozora.gr.jp/cards/000879/card43016.html)
genre	fiction
subcorpus	Aozora Bunko
terms of use	Creative Commons license

- 1 トロッコ
- 2 芥川龍之介
- 3 小田原 熱海 (あたみ) 間に、軽便鉄道敷設 (ふせつ) の工事が始まったのは、良平 (りょうへい) の八つの年だった。
- 4 良平は毎日村外 (はず) れへ、その工事を見物に行った。
- 5 工事を——といったところが、唯 (ただ) トロッコで土を運搬する——それが面白さに見に行ったのである。
- 6 トロッコの上には土工が二人、土を積んだ後 (うしろ) に作 (たたく) んでいる。
- 7 トロッコは山を下 (くだ) るのだから、人手を借りずに走って来る。
- 8 轆 (あお) るように車台が動いたり、土工の神天 (はんでん) の裾 (すそ) がひらついたり、細い線路がしなったり——良平はそいと思う事がある。
- 9 せめては一度でも土工と一緒に、トロッコへ乗りたいと思う事もある。
- 10 トロッコは村外れの平地へ来ると、自然と其站 (そこ) に止まってしまう。
- 11 と同時に土工たちは、身軽にトロッコを飛び降りるが早いから、その線路の終点へ車の土をぶちまける。
- 12 それから今度はトロッコを押し押し、もと来た山の方へ登り始める。
- 13 良平はその時乗れないまでも、押す事さえ出来たらと思うのである。
- 14 或 (ある) 夕方、——それは二月の初旬だった。
- 15 良平は二つ下の弟や、弟と同じ年の隣の子供と、トロッコの置いてある村外れへ行った。

8

Context view (コンテキスト表示) ページ

Toggle context view Go to Kusunoki Corpus overview

title	Torokko
author	Akutagawa, Ryunosuke
date	1922
source	Aozora Bunko (https://www.aozora.gr.jp/cards/000879/card43016.html)
genre	fiction
subcorpus	Aozora Bunko
terms of use	Creative Commons license

メタデータ

クリックで overview へ

- 1 トロッコ
- 2 芥川龍之介
- 3 小田原 熱海 (あたみ) 間に、軽便鉄道敷設 (ふせつ) の工事が始まったのは、良平 (りょうへい) の八つの年だった。
- 4 良平は毎日村外 (はず) れへ、その工事を見物に行った。
- 5 工事を——といったところが、唯 (ただ) トロッコで土を運搬する——それが面白さに見に行ったのである。
- 6 トロッコの上には土工が二人、土を積んだ後 (うしろ) に作 (たたく) んでいる。
- 7 トロッコは山を下 (くだ) るのだから、人手を借りずに走って来る。
- 8 轆 (あお) るように車台が動いたり、土工の神天 (はんでん) の裾 (すそ) がひらついたり、細い線路がしなったり——良平はそいと思う事がある。
- 9 せめては一度でも土工と一緒に、トロッコへ乗りたいと思う事もある。
- 10 トロッコは村外れの平地へ来ると、自然と其站 (そこ) に止まってしまう。
- 11 と同時に土工たちは、身軽にトロッコを飛び降りるが早いから、その線路の終点へ車の土をぶちまける。
- 12 それから今度はトロッコを押し押し、もと来た山の方へ登り始める。
- 13 良平はその時乗れないまでも、押す事さえ出来たらと思うのである。
- 14 或 (ある) 夕方、——それは二月の初旬だった。
- 15 良平は二つ下の弟や、弟と同じ年の隣の子供と、トロッコの置いてある村外れへ行った。

1行 = 1ツリー (文)

9

Context view (コンテキスト表示) ページ

Toggle context view Corpus overview

title	Aguni no Kami
author	Akutagawa, Ryunosuke
date	1921
source	Aozora Bunko (https://www.aozora.gr.jp/cards/000879/card43014.html)
genre	fiction
subcorpus	Aozora Bunko
terms of use	Creative Commons license

Lines:

Refresh selected lines

See analysis tree

ツリー番号をクリック → 分析を表示

- 1 アグニの首
- 2 芥川龍之介
- 3 一
- 4 支那の上海の或町です。
- 5 昼でも薄暗い或家の二階に、人相の悪い印度人の婆さんが一人、商人らしい一人の亜米利加人と何か頻に話し合
- 6 「実は今度もお婆さんに、古いを頼みに来たのだがね、——」
- 7 亜米利加人はそう言いながら、新しい巻煙草へ火をつけました。
- 8 「古いですか？」
- 9 古いは当分見ないことにしましたよ」
- 10 婆さんは囁るように、じろりと相手の顔を見ました。
- 11 「この頃は折角見て上げても、御礼さえ疎かにしない人が、多くなって来ましたからね」
- 12 「そりゃ勿論御礼をするよ」
- 13 亜米利加人は惜しげもなく、三百弗の小切手を一枚、婆さんの前へ投げてやりました。
- 14 「差当りこれだけ取って置けよ。」

10

Analysis view (分析表示) ページ

Context Previous Next free | source | tpt | newspan | indexing | flora | graph Corpus overview Query builder

↑

ID@Kainoki_7_aozora_Akutagawa-1921

亜米利加人はそう言いながら、新しい巻煙草へ火をつけました。

```

graph TD
    IP-MAT --- PP-SBJ[NP]
    IP-MAT --- IP-ADV-SCON[ADV]
    IP-MAT --- PU[、]
    IP-MAT --- PP[PP]
    IP-MAT --- PP-OB1[NP]
    IP-MAT --- VB[つけ]
    IP-MAT --- AX[まし]
    IP-MAT --- AXD[た]
    IP-MAT --- PU2[。]

    PP-SBJ --- NP1[NP]
    PP-SBJ --- P-OPTR[P-OPTR]
    NP1 --- N1[N]
    N1 --- 亜米利加人

    P-OPTR --- は

    IP-ADV-SCON --- ADV[そう]
    IP-ADV-SCON --- 言い
    IP-ADV-SCON --- ながら

    PP --- NP2[NP]
    PP --- P-ROLE[へ]
    NP2 --- IP-REL[IP-REL]
    IP-REL --- NP-SBJ[NP-SBJ]
    NP-SBJ --- ADJ[新しい]
    IP-REL --- N2[N]
    N2 --- 巻煙草

    PP-OB1 --- N3[N]
    N3 --- を

    PP-OB1 --- P-ROLE2[P-ROLE]
    P-ROLE2 --- 火
    
```

↓

Previous Next

11

セッション3の内容

1. ファイル一覧
2. KWIC 検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果のダウンロード
6. インターフェースに関するドキュメント

12

Overview から KWIC 検索へ

Toggle overview [tree](#) | [source](#) | [tftp](#) | [newspan](#) | [indexing](#) | [flora](#) | [graph](#) | [download](#) [Front page external link] [Help](#)

Tree search interface

Files:

TGrep2 search expression:

Search [basic](#) [Toggle tag set](#) [TGrep2 search guide](#)

487 files; 98,312 trees; 1,455,020 words

	Filename	tree count	word count	content description
1	aozora_Akutagawa-1921	289	4191	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2	aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3	aozora_Chiri-1955-7	248	9327	essay : Chiri, Mashiho (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4	aozora_Chiri-1956-2	42	1762	essay : Chiri, Mashiho (1956). <i>Ainugogaku</i>
5	aozora_Chiri-1956-7	30	966	essay : Chiri, Mashiho (1956). <i>Ainugo no Omoshirosa</i>
6	aozora_Chiri-1960-11	5	107	essay : Chiri, Mashiho (1960). "Aikokushin" <i>Watashi wa koo omou</i>
7	aozora_Dazai-1-1940	138	3035	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3
8	aozora_Dazai-2-1940	447	5037	fiction : Dazai, Osamu (1940). <i>Hashire Merusu</i> . Shincho, May edition
9	aozora_Doyle-1905	784	10533	fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Noouddo no ken</i>

13

Overview から KWIC 検索へ

Toggle overview [tree](#) | [source](#) | [tftp](#) | [newspan](#) | [indexing](#) | [flora](#) | [graph](#) | [download](#) [Front page external link] [Help](#)

Tree search interface

Files:

TGrep2 search expression:

Search [basic](#) [Toggle tag set](#) [TGrep2 search guide](#)

487 files; 98,312 trees; 1,455,020 words

**Toggle overview をクリック
→ Overview と Kwic 検索を循環**

	Filename	tree count	word count	content description
1	aozora_Akutagawa-1921	289	4191	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2	aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3	aozora_Chiri-1955-7	248	9327	essay : Chiri, Mashiho (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4	aozora_Chiri-1956-2	42	1762	essay : Chiri, Mashiho (1956). <i>Ainugogaku</i>
5	aozora_Chiri-1956-7	30	966	essay : Chiri, Mashiho (1956). <i>Ainugo no Omoshirosa</i>
6	aozora_Chiri-1960-11	5	107	essay : Chiri, Mashiho (1960). "Aikokushin" <i>Watashi wa koo omou</i>
7	aozora_Dazai-1-1940	138	3035	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3
8	aozora_Dazai-2-1940	447	5037	fiction : Dazai, Osamu (1940). <i>Hashire Merusu</i> . Shincho, May edition
9	aozora_Doyle-1905	784	10533	fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Noouddo no ken</i>

14

KWIC 検索

- KWIC 検索：テキストボックスに入力された文字列に機械的に一致するものを抽出する。
- 読点 (。) はテキストデータから除外されているので、検索文字列に使うことはできない。

Toggle overview [tree](#) | [source](#) | [indexed](#) | [span](#) | [tftp](#) | [flora2](#) | [collocates](#) | [graph](#) | [download](#)

String search interface

とこるを

Search [corpus order](#)

15

KWIC 検索

- 以下は、テキストボックスに検索文字列として「ところを」を入力したところ



KWIC 検索：表示順の種類

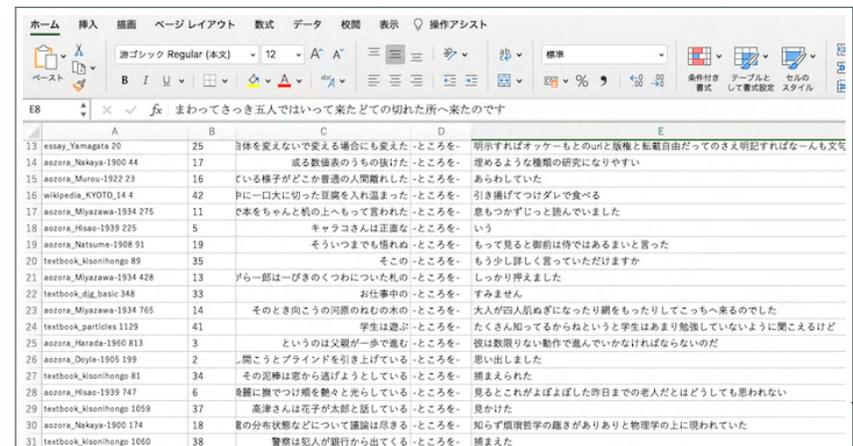
- corpus order (コーパス順 - デフォルト)
ファイル名のソート順と、ファイル中のツリー順に従った順番
- sort left (前文脈順)
検索文字列に先行する文字の順番
- sort right (後文脈順)
検索文字列に後続する文字の順番
- random
ランダムな順番

KWIC 検索：検索結果画面



KWIC 検索：検索結果の保存

- ブラウザのメニューから (あるいは右クリックして) 「すべてを選択」「コピー」 → 表計算ソフトやテキストエディタにペーストして保存



KWIC 検索

個々の用例の前後の文脈や解析情報を調べる

ファイル名の右側の番号 → Context view へ

チェックボックスの右側の番号 → Analysis view へ

KWIC 検索

個々の用例の前後の文脈や解析情報を調べる

Context view (コンテキスト表示) では、検索文字列を含む文がハイライトされるので、その前後の文脈を確認することができる

KWIC 検索

個々の用例の前後の文脈や解析情報を調べる

Analysis view (分析表示) では、検索文字列を含む文のツリーが図示されるので、解析情報を確認することができる

Context Previous Next Go to Kusumoki Corpus overview

page_182

ID@Kainoki_1051_textbook_kisonihongo

花子は太郎がその店に入るところを見かけた。

セッション3の内容

1. ファイル一覧
2. KWIC 検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果のダウンロード
6. インターフェースに関するドキュメント

ツリー検索

- ツリー内のノード（語・形態素の文字列、品詞タグ、句や節のタグ）やノード間の関係（前後関係や支配・非支配関係）を検索式で指定し、検索する。
- 本インターフェースは、TGrep2/Tregex を使った検索式に対応
- 検索式の作り方については session 4 と session 5 で扱う。
- 以下では、ファイルの指定、検索結果の表示方法、検索結果のダウンロードについて説明する。

24

(ツリー検索) 検索ファイルの指定

ツリー検索は、Overview ページから行う。

Toggle overview tree | source | ttp | newspan | indexing | flora | graph | download [Front page external link] Help

Tree search interface

Files:

TGrep2 search expression:

Search basic Toggle tag set TGrep2 search guide

487 files; 98,312 trees; 1,455,020 words

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4191	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9327	essay : Chiri, Mashihō (1955). <i>Ainu shuukyō seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1762	essay : Chiri, Mashihō (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	966	essay : Chiri, Mashihō (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashihō (1960). "Aikokushin" <i>Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3035	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3
8 aozora_Dazai-2-1940	447	5037	fiction : Dazai, Osamu (1940). <i>Hashire Merusu</i> . Shincho, May edition
9 aozora_Doyle-1905	784	10533	fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Noouddo no ken</i>

25

(ツリー検索) 検索ファイルの指定

ツリー検索は、Overview ページから行う。

Toggle overview tree | source | ttp | newspan | indexing | flora | graph | download [Front page external link] Help

Tree search interface

Files:

TGrep2 search expression:

Search basic Toggle tag set TGrep2 search guide

487 files; 98,312 trees; 1,455,020 words

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4191	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9327	essay : Chiri, Mashihō (1955). <i>Ainu shuukyō seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1762	essay : Chiri, Mashihō (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	966	essay : Chiri, Mashihō (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashihō (1960). "Aikokushin" <i>Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3035	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3
8 aozora_Dazai-2-1940	447	5037	fiction : Dazai, Osamu (1940). <i>Hashire Merusu</i> . Shincho, May edition
9 aozora_Doyle-1905	784	10533	fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Noouddo no ken</i>

検索対象ファイルを指定
(何も指定しないと全ファイルが対象になる)
「ファイル番号」または「ファイル名」で指定する

26

(ツリー検索) 検索ファイルの指定

ファイル番号で指定

- 1 番目のファイルだけ
Files: 1p
- 1 番目と 10 番目のファイル
Files: 1p;10p
- 1 番目から 10 番目までのファイル
Files: 1,10p
- 1 から 10 番目までと、15 から 20 番目までのファイル
Files: 1,10p;15,20p

27

(ツリー検索) 検索ファイルの指定

ファイル名で指定

- wikipedia_Kamen_Rider だけ
(ファイル名を半角スラッシュで囲む)

```
Files: /wikipedia_Kamen_Rider/
```

- aozora_Akutagawa-1921 と
aozora_Akutagawa-1922
(ファイル名の間には \ | を入れ、全体を半角スラッシュで囲む)

```
Files: /aozora_Akutagawa-1921\|aozora_Akutagawa-1922/
```

28

(ツリー検索) 検索ファイルの指定

検索ファイルを指定し、検索式を入力しない状態で、[Search] をクリックすると、検索対象ファイル全体のツリー数と語数が表示される。

① ファイルを指定

② クリック

③ 指定したファイルのツリー数と語数

Filename	tree count	word count	content description
66 bible_nt	1025	15238	sample : The New Testament (trans.)
67 bible_ot	639	10851	sample : The Old Testament (trans.)

29

セッション3の内容

1. ファイル一覧
2. KWIC 検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果のダウンロード
6. インターフェースに関するドキュメント

30

(ツリー検索) 検索結果の表示方法

検索式を入力して [Search] をクリックすると、検索が実行される。

検索式を入力

120 files; 20,249 trees; 355,372 words

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4191	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9327	essay : Chiri, Mashiho (1955). <i>Ainu shuukyoo seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1762	essay : Chiri, Mashiho (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	966	essay : Chiri, Mashiho (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashiho (1960). "Aikokushin" <i>Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3035	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3

31

(ツリー検索) 検索結果の表示方法

検索式を入力して [Search] をクリックすると、検索を実行する。

- 検索結果の表示画面の上部に、検索式にヒットした数とヒット含むファイルの数が示される。
- 検索対象ファイルを指定していれば、頻度情報(10,000語あたりのヒット数)も示される。

32

(ツリー検索) 検索結果の表示方法

3つの表示方法:

basic (デフォルト) と graphical と brackets

The screenshot shows the 'Tree search interface' with a search expression 'PASS' and a dropdown menu for 'display method' (表示方法の選択) with options: basic, graphical, and brackets. Below the interface is a table of search results:

Filename	tree count	word count	content description
1 aozora_Akutagawa-1921	289	4191	fiction : Akutagawa, Ryunosuke (1921). <i>Aguni no Kami</i>
2 aozora_Akutagawa-1922	152	2378	fiction : Akutagawa, Ryunosuke (1922). <i>Torokko</i>
3 aozora_Chiri-1955-7	248	9327	essay : Chiri, Mashiho (1955). <i>Ainu shuuyoo seiritsu no shiteki haikai</i>
4 aozora_Chiri-1956-2	42	1762	essay : Chiri, Mashiho (1956). <i>Ainugogaku</i>
5 aozora_Chiri-1956-7	30	966	essay : Chiri, Mashiho (1956). <i>Ainugo no Omoshiroa</i>
6 aozora_Chiri-1960-11	5	107	essay : Chiri, Mashiho (1960). "Aikokushin" <i>Watashi wa koo omou</i>
7 aozora_Dazai-1-1940	138	3035	essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3

33

(ツリー検索) 検索結果の表示方法

basic

- 文を表示 (表示数の上限なし)
- 検索式のマスターノードにマッチした部分がハイライトされる。

The screenshot shows search results for the pattern 'PASS'. The text is displayed with the search results highlighted in yellow. The first few results are:

- 36 aozora_Akutagawa-1921
惠蓮はいくら叱られても、じっと俯向いたまま黙っていました。
- 115 aozora_Akutagawa-1921
が、その拍子に婆さんが、鴉の啼くような声を立てたかと思うと、まるで電気に打たれたように、ピストル
- 121 aozora_Akutagawa-1921
遠藤はとうとうたまり兼ねて、火花の旋風に追われながら、転げるように外へ逃げ出しました。
- 156 aozora_Akutagawa-1921
婆さんの前には心配そうな惠蓮が、——いや、支那服を着せられた妙子が、じっと椅子に坐っていました。
- 160 aozora_Akutagawa-1921
しかし今うっかりそんな気が、婆さんの眼にでも止まったが最後、この恐い魔法使いの家から、逃げ出す
- 206 aozora_Akutagawa-1921
私はまだお前に敗される程、毫縁はしていない心算だよ。
- 277 aozora_Akutagawa-1921

34

(ツリー検索) 検索結果の表示方法

graphical

- グラフィカルなツリー (50の検索結果まで、表示内容は検索の実行毎にランダムに選ばれる)
- 検索式のマスターノードにマッチした部分がハイライトされる。

The screenshot shows search results for the pattern 'PASS' in graphical view. It displays a tree diagram for the sentence: 瓦筋屋根が笑いて、電燈が附けてある。 The tree structure is as follows:

```

    graph TD
      IPMAT[IPMAT] --- IPADV[IP-ADV]
      IPMAT --- PU1[PU]
      IPMAT --- PP[PP]
      IPMAT --- VB[VB]
      IPMAT --- PC[PCONN]
      IPMAT --- PASS[PASS]
      IPMAT --- VB2[VB2]
      IPMAT --- PU2[PU]
      IPADV --- PP[PP]
      IPADV --- VB[VB]
      IPADV --- PC[PCONN]
      PP --- NP1[NP]
      PP --- PROLE[PROLE]
      NP1 --- N1[N]
      NP1 --- BT1[BT]
      PROLE --- BT2[BT]
      VB --- 笑い[笑い]
      PC --- て[て]
      VB2 --- 電燈[電燈]
      PU1 --- 瓦筋屋根[瓦筋屋根]
      PU2 --- 電燈[電燈]
  
```

35

(ツリー検索) 検索結果の表示方法

brackets

- 文と括弧付きツリー (表示数の上限なし)
- 検索式のマスターノードにマッチした部分がハイライトされる。

Search pattern: PASS
The search returned 1759 hits. 120 different texts were searched (355368 words [120 texts]; frequency of hits: 49.50 instances per ten thousand words). Hits

Download all results comma-separated values

See analysis tree

思慮はいくら叱られても、じっと俯向いたまま黙っていました。

1 ((IP-MAT (PP-SBJ (NP (NPR 思慮)) (P-OPTR は)) (PP-SCON-OND (IP-ADV (NP-LGS srow) (ADVP (WADV いくら)) (VB 叱ら) (PASS け) た)) (N さま)) (VP 黙っ) (P-COMP て) (VBJ い) (AX まし) (AX た) (PU .)) (ID 36_mozora_Akutaqawa-1921;1P1)

2 PRD (IP-EMB (PP-SBJ srow) (P-ROLE の)) (VB 鳴<)) (N よう)) (AX な) (N 声)) (P-ROLE を)) (VB 立て) (AX た) (P-FINAL か (IP-EMB (NP-SBJ srow) (PP-LGS (NP (N 電報)) (P-ROLE に)) (VB 打た) (PASS け) (AX た)) (N よう)) (AX に)) (PU .)) (PP-SBJ の) (AX まし) (AX た) (PU .)) (ID 135_mozora_Akutaqawa-1921;1P1)

3 ((IP-MAT (PP-SBJ (NP (NPR 思慮)) (P-OPTR は)) (IP-ADV-CONJ (ADVP (ADV どうとう)) (VB たまひ) (VBJ 黙ね) (P-COMP て) (PU .)) (PASS け) (P-COMP ながら)) (PU .)) (PP (NP (IP-EMB (VB 恥げる)) (N よう)) (P-ROLE に)) (PP (NP (N 男)) (P-ROLE へ)) (VB 逃げ出 皆さんの前には心配そうな思慮が、——い中、支那軍を管轄された砂子が、じっと椅子に坐っていました。

4 ((IP-MAT (PP (NP (PP (NP (N 妻さん)) (P-ROLE の)) (N 前)) (P-ROLE に)) (P-OPTR は)) (PRN (PP (NP (IP-REL (NP-SBJ ち) (ADJN 4 SBJ (NP (IP-REL (NP-SBJ ち) (NP-LGS srow) (PP-OB1 (NP (N 支那軍)) (P-ROLE を)) (VB 撃せ) (PASS られ) (AX た)) (NPR 砂子) (COMP て) (VBJ い) (AX まし) (AX た) (PU .)) (ID 156_mozora_Akutaqawa-1921;1P1)

5 ((IP-MAT (NP-LGS srow) (CONJ かし) (IP-ADV-SCON (NP-SBJ srow) (PP-SCON-OND (IP-ADV (NP-TMP (N 今)) (ADVP (ADV うっかり)) (P-ROLE の)) (N 間)) (P-ROLE に)) (P-OPTR でも)) (VB 止まっ) (AX た) (P-COMP の)) (NP-PRD (N 思慮)) (AX ち) (PU .)) (PP-SBJ (NP (しい) (N 魔法使い)) (P-ROLE の)) (N 間)) (P-ROLE から)) (PU .)) (VB 逃げ出せ) (AX う) (P-COMP という) (N 間)) (P-OPTR は)) (PU (ND て) (PU .)) (ID 108_mozora_Akutaqawa-1921;1P1)

36

(ツリー検索) 検索結果の表示方法

注意:

basic と brackets では、ひとつのツリーの中に複数のヒットがあっても、そのツリーを1度しか表示しない。

- 例えば、検索式「ところ」で検索したときに、以下のツリー (文) のようにヒットが2つあっても、画面に同じツリーが2度表示されることはない。

というのは、どう考えても、彼自身が何かしら憑きものがあるような、よく町裏の小暗いところを歩いていたりしている様子が、どこか普通の人間離れしたところをあらわしていた。

- ヒット数と表示されたツリーの数が違うことがありうる。 37

セッション3の内容

1. ファイル一覧
2. KWIC 検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果のダウンロード
6. インターフェースに関するドキュメント

38

(ツリー検索) 検索結果のダウンロード

- 結果表示画面の上部の [Download all results] をクリック
- 2種類の形式: CSV形式 / 括弧付きツリー形式

39

(ツリー検索) 検索結果のダウンロード

CSV

"1402";"597_bible_nt;b.JOH.7.47;JP" 「あなたがたまでが、
だまさ <mark>れ</mark> ているのではないか。”
"1537";"2_book_excerpt-51;JP" なぜ 翼 は 失わ <mark>れ
</mark> た の か”

- “ 番号 (検索結果中の通し番号) ”, “ ファイルと
ファイル中の番号 ”, “ 例文 ”
- 例文中のマスターノードとのマッチは
<mark>...</mark> で示される。

40

(ツリー検索) 検索結果のダウンロード

括弧付きツリー

((CP-FINAL (PUL っ) (IP-SUB (NP-LGS *pro*) (PP-SBJ (NP;POLICE (PRO あ
なたがた)) (P-OPTR まで) (P-ROLE が)) (PU、) (VB だまさ) (__PASS れ)
(P-CONN て) (VB2 いる) (FN の) (AX で) (P-OPTR は) (NEG ない) (P-FINAL
か) (PU。)) (ID 597_bible_nt;b.JOH.7.47;JP))
((CP-QUE (IP-SUB (NP-LGS *pro*) (ADVP (WADV なぜ)) (PP-SBJ
(NP;WINGS_TO_HEAVEN (N 翼)) (P-OPTR は) (VB 失わ) (__PASS れ) (AXD
た) (FN の)) (P-FINAL か)) (ID 2_book_excerpt-51;JP))

- 例文中のマスターノードとのマッチは “__” で示
される。

41

(ツリー検索) 検索結果のダウンロード

注意:

- CSV でも 括弧付きツリーでも、ひとつのツリー
に複数のヒットがあっても、そのツリーはファ
イル中に1つしか含まれない (1文中に複数回
“<mark>...</mark>” あるいは “__” が含まれ
うる)。
- 括弧付きツリー形式のファイルは、専用のツ
ール (例えば、Tregex ([https://nlp.stan-
ford.edu/software/tregex.shtml](https://nlp.stanford.edu/software/tregex.shtml))) をローカル
にインストールして利用するとよい。

42

セッション3の内容

1. ファイル一覧
2. KWIC 検索
3. (ツリー検索) ファイルの指定
4. (ツリー検索) 検索結果の表示方法
5. (ツリー検索) 検索結果のダウンロード
6. インターフェースに関するドキュメント

43

インターフェースに関するドキュメント

The screenshot shows a web interface for a text search tool. At the top, there are navigation links: [tree](#) | [source](#) | [indexed](#) | [span](#) | [tptp](#) | [flora2](#) | [collocates](#) | [graph](#) | [download](#) | [Main page](#) | [Help](#). Below this is a section titled "Tree search interface" with a "Files:" input field and a "Tregex search expression:" input field. There are three callout boxes with orange borders and arrows pointing to specific elements:

- A callout box at the top right points to the "Help" link with the text: **Help : インターフェースガイド置き場へ**
- A callout box at the bottom left points to the "Toggle tag set" link with the text: **Toggle tag set : タグ一覧の表示・非表示**
- A callout box at the bottom right points to the "Tregex search guide" link with the text: **Tregex search guide : Tregex のガイド置き場へ**

Below the search fields is a table with the following columns: tree, word count, content description. The table contains the following data:

tree	word count	content description
1	89	4192 fiction : Akutaga
2	52	2378 fiction : Akutaga
3	48	9308 essay : Chiri, Ma
4	42	1761 essay : Chiri, Ma
5	30	967 essay : Chiri, Mashihō (1956). <i>Ainugo no Omoshirosa</i>
6	5	107 essay : Chiri, Mashihō (1960). <i>"Aikokushin" Watashi wa koo omou</i>
7	138	3044 essay : Dazai, Osamu (1940). <i>Sake-girai</i> . Chisei Vol. 3, No. 3
8	447	5043 fiction : Dazai, Osamu (1940). <i>Hashire Merosu</i> . Shincho, May edition
9	781	10538 fiction : Doyle, Arthur Conan — Kareha (trans.) (2000). <i>Nooudo no kenchikuka</i>
10	306	9709 fiction : Edogawa, Ranpo (1929). <i>Oshie to tabi suru otoko</i> . Shinseinen, June edition
11	1232	29115 fiction : Kafka, Franz (translator: Harada, Yoshito) (1915 (1960)). <i>Die Verwandlung (Henshin)</i> . Kafka Zenshu