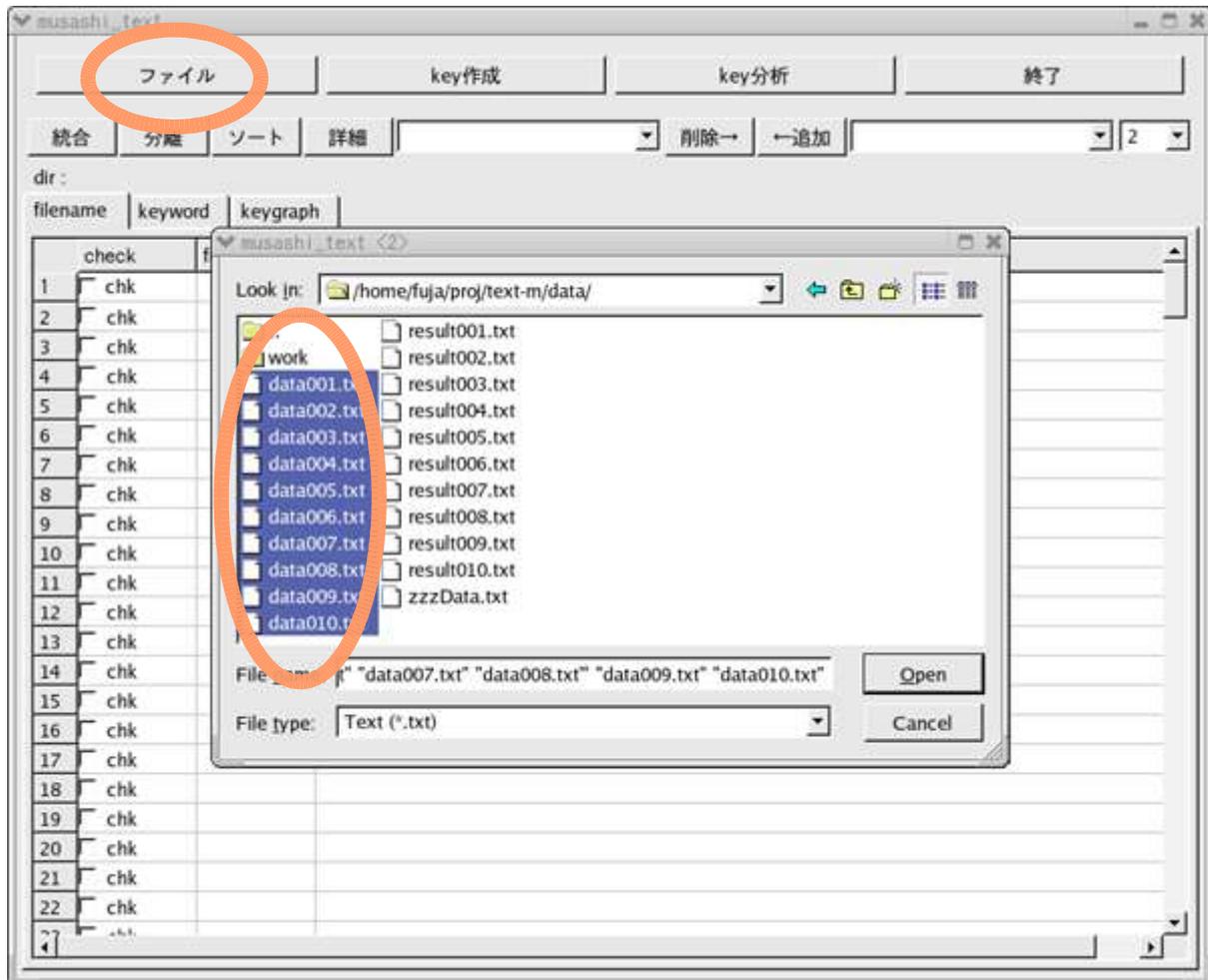


1. ファイルの指定



ファイルボタンを押しダイアログを出し、分析対象となるテキストファイルを指定します。現状ではテキストファイルの日本語コードに `eut-jp` を想定しています。必要であれば予め `eut-jp` へのコード変換をお願いします。

3. キーワード候補の調整

3. 1. ソーティング

| check | index | count | word | kana | base | part | reserve1 | rese |
|-------|-------|-------|-------|-------|-------|-----------|----------|------|
| 1 | chk | 2 | の | ノ | | 助詞-格助詞-一般 | | |
| 2 | chk | 7 | に | ニ | に | 助詞-格助詞-一般 | | |
| 3 | chk | 8 | は | ハ | は | 助詞-係助詞 | | |
| 4 | chk | 9 | , | , | , | 記号-読点 | | |
| 5 | chk | 39 | が | ガ | が | 助詞-格助詞-一般 | | |
| 6 | chk | 36 | . | . | . | 記号-句点 | | |
| 7 | chk | 144 | なる | ナル | なる | 動詞-自立 | 五段・ラ行 | 基本 |
| 8 | chk | 4 | を | ヲ | を | 助詞-格助詞-一般 | | |
| 9 | chk | 48 | し | シ | する | 動詞-自立 | サ変・スル | 連用 |
| 10 | chk | 35 | ます | マス | ます | 助動詞 | 特殊・マス | 基本 |
| 11 | chk | 124 | フィルター | フィルター | フィルター | 名詞-一般 | | |
| 12 | chk | 135 | と | ト | と | 助詞-並立助詞 | | |
| 13 | chk | 81 | で | デ | だ | 助動詞 | 特殊・ダ | 連用 |
| 14 | chk | 119 | で | デ | で | 助詞-格助詞-一般 | | |
| 15 | chk | 56 | 0 | ゼロ | 0 | 名詞-数 | | |
| 16 | chk | 94 | て | テ | て | 助詞-接続助詞 | | |
| 17 | chk | 1 | 信号 | シンゴウ | 信号 | 名詞-一般 | | |
| 18 | chk | 18 | 特性 | トクセイ | 特性 | 名詞-一般 | | |
| 19 | chk | 128 | です | デス | です | 助動詞 | 特殊・デス | 基本 |
| 20 | chk | 17 | 伝達 | デンタツ | 伝達 | 名詞-サ変接続 | | |
| 21 | chk | 111 | こと | コト | こと | 名詞-非自立-一般 | | |
| 22 | chk | 777 | . | . | . | 記号-一般 | | |

茶筌が生成したキーワード候補には固有な index 番号が振られ上に示すようにテーブルに並べられます。テーブルにおいて word 項には元の文章にあった単語そのまま、base 項には語句の変形を考慮して得られる単語が並べられます。base 項を元に重複をチェックし同じキーワード候補については個数を数え count 項に納めます。内部的には、ここで同じと判定されたキーワードについて元の文章中での出現パターンを使ってテキストマイニングを行います。出現パターンが似ているキーワード同士がもしあれば、おおむね関連性が強いと判定されるでしょう。

キーワード候補の様子を確認し調整する目的で、index 項、count 項、base 項、part (品詞) 項を基準にしてソーティングができるようになっています。上記いずれかの項目にマウスを合わせクリックにより選択した後に、ソートボタンを押してください。

3. 2. 類義語の統合

The screenshot shows the 'musashi_text' application window. At the top, there are tabs for 'ファイル', 'key作成', 'key分析', and '終了'. Below these are buttons for '統合', '分離', 'ソート', '詳細', a dropdown menu for '名詞-サ変接続', '削除→', '←追加', another dropdown for '副詞一般', and a numeric input '5'. The main area contains a table with columns: 'check', 'index', 'count', 'word', 'kana', 'base', 'part', 'reserve1', and 'rese'. The table lists various keywords with their corresponding counts and parts of speech. Two checkboxes in the 'check' column are highlighted with orange circles.

| | check | index | count | word | kana | base | part | reserve1 | rese |
|----|-------------------------------------|-------|-------|----------|--------|-------|-----------|-----------|------|
| 1 | <input type="checkbox"/> | chk | 124 | 32 フィルター | フィルター | フィルター | 名詞一般 | | |
| 2 | <input type="checkbox"/> | chk | 1 | 23 信号 | シゴウ | 信号 | 名詞一般 | | |
| 3 | <input checked="" type="checkbox"/> | chk | 18 | 21 特性 | トクセイ | 特性 | 名詞一般 | | |
| 4 | <input type="checkbox"/> | chk | 17 | 19 伝達 | デンタツ | 伝達 | 名詞-サ変接続 | | |
| 5 | <input checked="" type="checkbox"/> | chk | 25 | 15 帯域 | タイイキ | 帯域 | 名詞一般 | | |
| 6 | <input type="checkbox"/> | chk | 107 | 15 回復 | カイフク | 回復 | 名詞-サ変接続 | | |
| 7 | <input type="checkbox"/> | chk | 62 | 12 逆 | ギャク | 逆 | 名詞一般 | | |
| 8 | <input type="checkbox"/> | chk | 438 | 11 ノイズ | ノイズ | ノイズ | 名詞一般 | | |
| 9 | <input type="checkbox"/> | chk | 122 | 10 的 | テキ | 的 | 名詞-接尾-形容動 | | |
| 10 | <input type="checkbox"/> | chk | 649 | 10 大きく | オオキク | 大きい | 形容詞-自立 | 形容詞・イ段 | 連用 |
| 11 | <input type="checkbox"/> | chk | 12 | 8 上 | ウエ | 上 | 名詞一般 | | |
| 12 | <input type="checkbox"/> | chk | 379 | 8 観測 | カンソク | 観測 | 名詞-サ変接続 | | |
| 13 | <input type="checkbox"/> | chk | 22 | 7 周波数 | シュウハスウ | 周波数 | 名詞一般 | | |
| 14 | <input type="checkbox"/> | chk | 108 | 7 処理 | ショリ | 処理 | 名詞-サ変接続 | | |
| 15 | <input type="checkbox"/> | chk | 1245 | 7 カーネル | カーネル | カーネル | 名詞一般 | | |
| 16 | <input type="checkbox"/> | chk | 560 | 6 良い | ヨイ | 良い | 形容詞-自立 | 形容詞・アウオ基本 | |
| 17 | <input type="checkbox"/> | chk | 676 | 6 増幅 | ゾウフク | 増幅 | 名詞-サ変接続 | | |
| 18 | <input type="checkbox"/> | chk | 589 | 5 小さい | チイサイ | 小さい | 形容詞-自立 | 形容詞・アウオ基本 | |
| 19 | <input type="checkbox"/> | chk | 14 | 4 図 | ズ | 図 | 名詞一般 | | |
| 20 | <input type="checkbox"/> | chk | 182 | 4 値 | ネ | 値 | 名詞一般 | | |
| 21 | <input type="checkbox"/> | chk | 246 | 4 計算 | ケイサン | 計算 | 名詞-サ変接続 | | |
| 22 | <input type="checkbox"/> | chk | 608 | 4 場合 | バイ | 場合 | 名詞-副詞可能 | | |

base 項を元にした比較では同じキーワードであると判断されない場合にも、同じものとして扱いたい場合があります。そのようなキーワードを以下の操作によりまとめることができます。まとめたキーワードの左にあるボックスにチェックを入れ、上の統合ボタンを押します。チェックが入っていたキーワードは count 項の数値が大きいものにまとめられます。

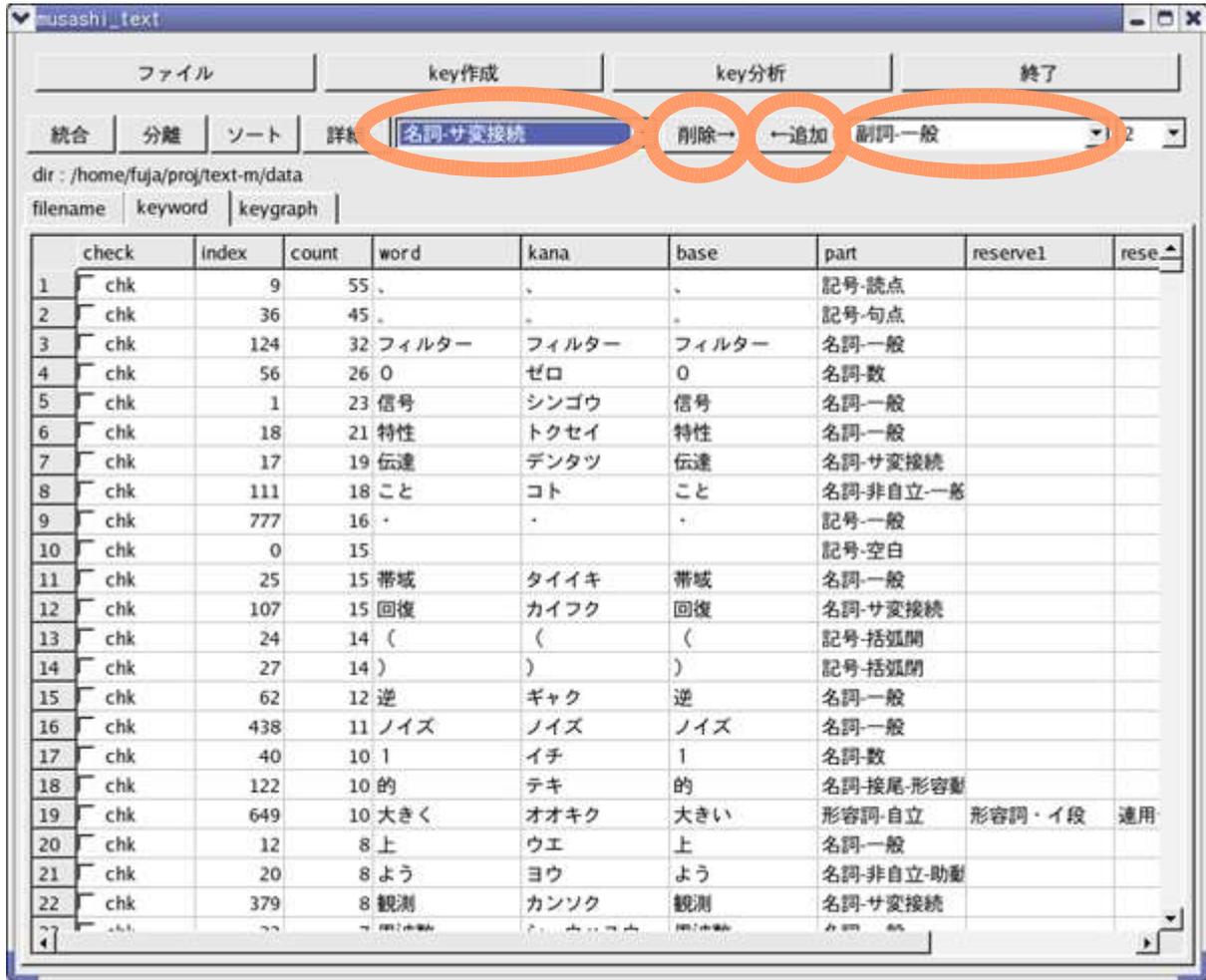
3. 3. 類義語の分離



| | check | Index | count | word | kana | base | part | reserve1 | rese |
|----|-------------------------------------|-------|-------|----------|--------|-------|-----------|------------|------|
| 1 | <input type="checkbox"/> | chk | 124 | 32 フィルター | フィルター | フィルター | 名詞-一般 | | |
| 2 | <input type="checkbox"/> | chk | 1 | 23 信号 | シゴウ | 信号 | 名詞-一般 | | |
| 3 | <input checked="" type="checkbox"/> | chk | 18 | 36 特性 | トクセイ | 特性 | 名詞-一般 | | |
| 4 | <input type="checkbox"/> | chk | 17 | 19 伝達 | デンタツ | 伝達 | 名詞-サ変接続 | | |
| 5 | <input type="checkbox"/> | chk | 107 | 15 回復 | カイフク | 回復 | 名詞-サ変接続 | | |
| 6 | <input type="checkbox"/> | chk | 62 | 12 逆 | ギャク | 逆 | 名詞-一般 | | |
| 7 | <input type="checkbox"/> | chk | 438 | 11 ノイズ | ノイズ | ノイズ | 名詞-一般 | | |
| 8 | <input type="checkbox"/> | chk | 122 | 10 的 | テキ | 的 | 名詞-接尾-形容動 | | |
| 9 | <input type="checkbox"/> | chk | 649 | 10 大きく | オオキク | 大きい | 形容詞-自立 | 形容詞・イ段 | 連用 |
| 10 | <input type="checkbox"/> | chk | 12 | 8 上 | ウエ | 上 | 名詞-一般 | | |
| 11 | <input type="checkbox"/> | chk | 379 | 8 観測 | カンソク | 観測 | 名詞-サ変接続 | | |
| 12 | <input type="checkbox"/> | chk | 22 | 7 周波数 | シュウハスウ | 周波数 | 名詞-一般 | | |
| 13 | <input type="checkbox"/> | chk | 108 | 7 処理 | ショリ | 処理 | 名詞-サ変接続 | | |
| 14 | <input type="checkbox"/> | chk | 1245 | 7 カーネル | カーネル | カーネル | 名詞-一般 | | |
| 15 | <input type="checkbox"/> | chk | 560 | 6 良い | ヨイ | 良い | 形容詞-自立 | 形容詞・アウオ段基本 | |
| 16 | <input type="checkbox"/> | chk | 676 | 6 増幅 | ゾウフク | 増幅 | 名詞-サ変接続 | | |
| 17 | <input type="checkbox"/> | chk | 589 | 5 小さい | チイサイ | 小さい | 形容詞-自立 | 形容詞・アウオ段基本 | |
| 18 | <input type="checkbox"/> | chk | 14 | 4 図 | ズ | 図 | 名詞-一般 | | |
| 19 | <input type="checkbox"/> | chk | 182 | 4 値 | ネ | 値 | 名詞-一般 | | |
| 20 | <input type="checkbox"/> | chk | 246 | 4 計算 | ケイサン | 計算 | 名詞-サ変接続 | | |
| 21 | <input type="checkbox"/> | chk | 608 | 4 場合 | バアイ | 場合 | 名詞-副詞可能 | | |
| 22 | <input type="checkbox"/> | chk | 812 | 4 パワー | パワー | パワー | 名詞-一般 | | |

分離したいキーワードの左にあるボックスにチェックを入れ、上の分離ボタンを押します。base 項の比較では同一と判定されないキーワードについては分離されません。

3. 4. 品詞による選択



除外したい品詞については、左のリストで指定を行い削除ボタンを押してください。該当する品詞のキーワード候補がテーブルから除外されます。
追加したい品詞については、右のリストで指定を行い追加ボタンを押してください。該当する品詞のキーワード候補がテーブルに追加されます。

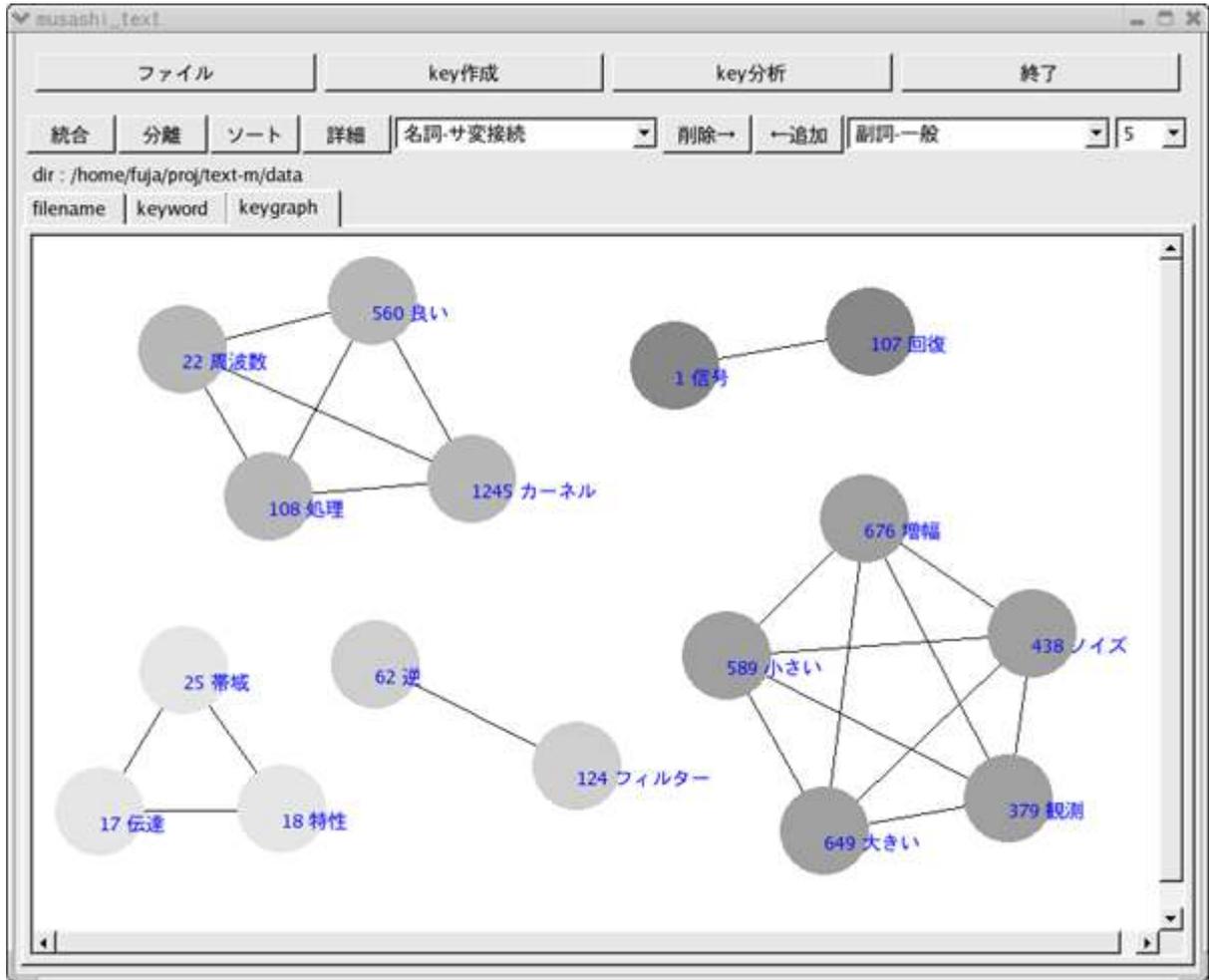
7. キーワードの分析

The screenshot shows a software window titled 'musashi_text'. It has a menu bar with 'ファイル', 'key作成', 'key分析', and '終了'. Below the menu bar are buttons for '統合', '分離', 'ソート', '詳細', a dropdown menu set to '名詞-サ変接続', '削除→', '←追加', and '副詞-一般'. A dropdown menu on the far right shows the number '4'. The main area has a directory path 'dir: /home/fuja/proj/text-m/data' and tabs for 'filename', 'keyword', and 'keygraph'. A table with the following columns is displayed: 'checked', 'index', 'count', 'word', 'kana', 'base', 'part', 'reserve1', and 'reserve2'. The 'checked' column contains checkboxes, with a vertical oval highlighting this column. The table contains 22 rows of data.

| checked | index | count | word | kana | base | part | reserve1 | reserve2 |
|-------------------------------------|-------|-------|-------|--------|-------|-----------|------------|----------|
| <input checked="" type="checkbox"/> | 124 | 32 | フィルター | フィルター | フィルター | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 1 | 23 | 信号 | シゴウ | 信号 | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 18 | 21 | 特性 | トクセイ | 特性 | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 17 | 19 | 伝達 | デンタツ | 伝達 | 名詞-サ変接続 | | |
| <input checked="" type="checkbox"/> | 25 | 15 | 帯域 | タイキ | 帯域 | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 107 | 15 | 回復 | カイフク | 回復 | 名詞-サ変接続 | | |
| <input checked="" type="checkbox"/> | 62 | 12 | 逆 | ギャク | 逆 | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 438 | 11 | ノイズ | ノイズ | ノイズ | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 122 | 10 | 的 | テキ | 的 | 名詞-接尾-形容動 | | |
| <input checked="" type="checkbox"/> | 649 | 10 | 大きく | オオキク | 大きい | 形容詞-自立 | 形容詞・イ段 | 連用 |
| <input checked="" type="checkbox"/> | 12 | 8 | 上 | ウエ | 上 | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 379 | 8 | 観測 | カンソク | 観測 | 名詞-サ変接続 | | |
| <input checked="" type="checkbox"/> | 22 | 7 | 周波数 | シュウハスウ | 周波数 | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 108 | 7 | 処理 | ショリ | 処理 | 名詞-サ変接続 | | |
| <input checked="" type="checkbox"/> | 1245 | 7 | カーネル | カーネル | カーネル | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 560 | 6 | 良い | ヨイ | 良い | 形容詞-自立 | 形容詞・アウオ段基本 | |
| <input checked="" type="checkbox"/> | 676 | 6 | 増幅 | ゾウフク | 増幅 | 名詞-サ変接続 | | |
| <input checked="" type="checkbox"/> | 589 | 5 | 小さい | チイサイ | 小さい | 形容詞-自立 | 形容詞・アウオ段基本 | |
| <input checked="" type="checkbox"/> | 14 | 4 | 図 | ズ | 図 | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 182 | 4 | 値 | ネ | 値 | 名詞-一般 | | |
| <input checked="" type="checkbox"/> | 246 | 4 | 計算 | ケイサン | 計算 | 名詞-サ変接続 | | |
| <input checked="" type="checkbox"/> | 608 | 4 | 場合 | バイ | 場合 | 名詞-副詞可能 | | |

調整が終わったキーワード候補の中でテキストマイニングの対象としたいものについては左のボックスにチェックを入れてください。右のボックスにはテキストマイニングの際に想定するクラスター数（分類数）を指定してください。key分析ボタンを押すとテキストマイニングが始まります。

8. 分析結果



テキストマイニングの結果は以上のように表示されます。

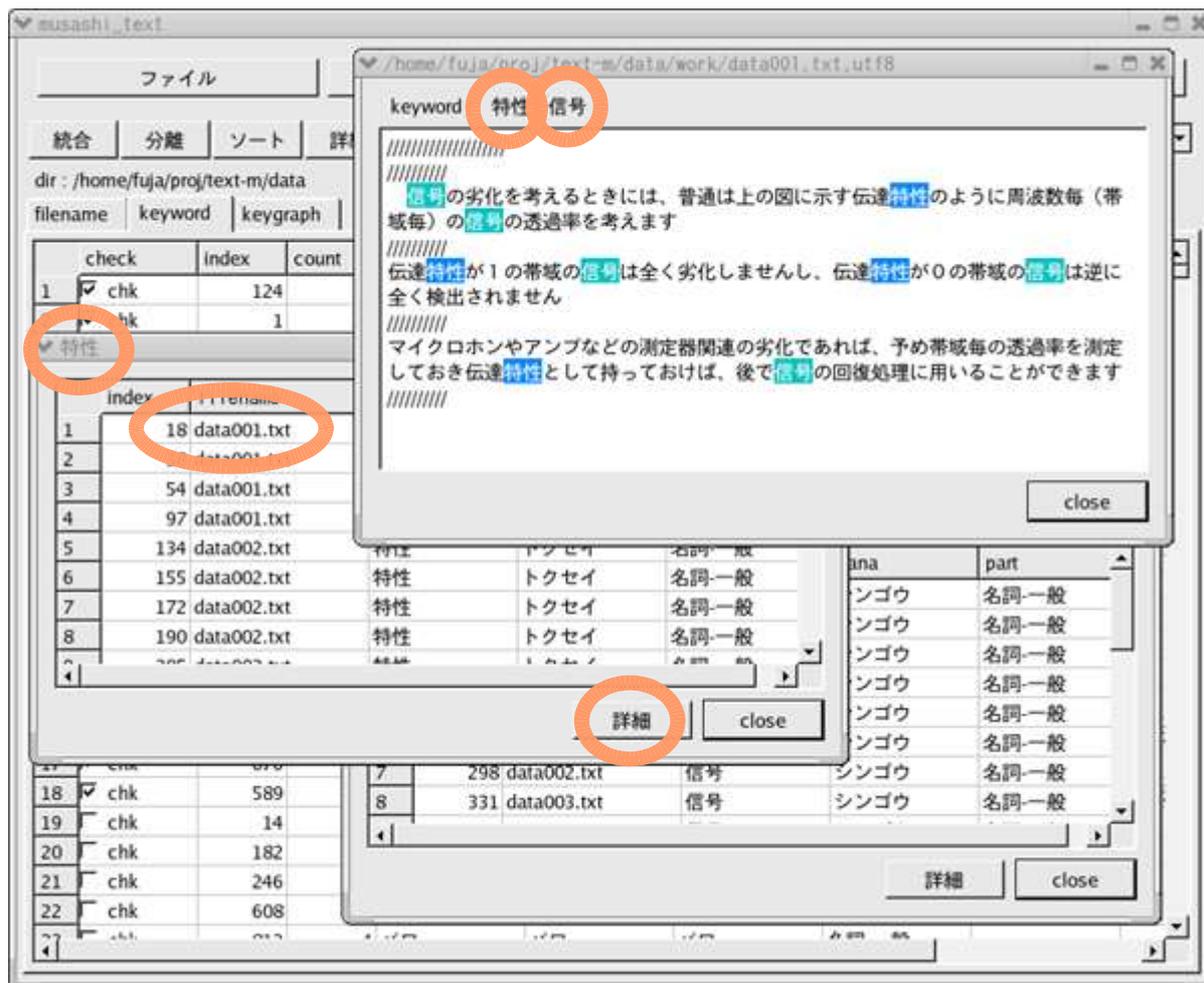
9. キーワードの詳細表示

| check | index | count | word | kana | base | part | reservel | rese |
|-------|-------|-------|----------|-------|-------|---------|----------|------|
| 1 | chk | 124 | 32 フィルター | フィルター | フィルター | 名詞一般 | | |
| 2 | chk | 1 | 23 信号 | シゴウ | 信号 | 名詞一般 | | |
| 3 | chk | 18 | 21 特性 | トクセイ | 特性 | 名詞一般 | | |
| 4 | chk | 17 | 19 伝達 | デンタツ | 伝達 | 名詞-サ変接続 | | |
| 5 | chk | 25 | 25 帯域 | タイキ | 帯域 | 名詞一般 | | |
| 6 | chk | 107 | 107 信号 | シゴウ | 信号 | 名詞一般 | | |
| 7 | chk | 62 | 62 信号 | シゴウ | 信号 | 名詞一般 | | |
| 8 | chk | 438 | 438 信号 | シゴウ | 信号 | 名詞一般 | | |
| 9 | chk | 122 | 122 信号 | シゴウ | 信号 | 名詞一般 | | |
| 10 | chk | 649 | 649 信号 | シゴウ | 信号 | 名詞一般 | | |
| 11 | chk | 12 | 12 信号 | シゴウ | 信号 | 名詞一般 | | |
| 12 | chk | 379 | 379 信号 | シゴウ | 信号 | 名詞一般 | | |
| 13 | chk | 22 | 22 信号 | シゴウ | 信号 | 名詞一般 | | |
| 14 | chk | 108 | 108 信号 | シゴウ | 信号 | 名詞一般 | | |
| 15 | chk | 1245 | 1245 信号 | シゴウ | 信号 | 名詞一般 | | |
| 16 | chk | 560 | 560 信号 | シゴウ | 信号 | 名詞一般 | | |
| 17 | chk | 676 | 676 信号 | シゴウ | 信号 | 名詞一般 | | |
| 18 | chk | 589 | 589 信号 | シゴウ | 信号 | 名詞一般 | | |
| 19 | chk | 14 | 14 信号 | シゴウ | 信号 | 名詞一般 | | |
| 20 | chk | 182 | 182 信号 | シゴウ | 信号 | 名詞一般 | | |
| 21 | chk | 246 | 246 信号 | シゴウ | 信号 | 名詞一般 | | |
| 22 | chk | 608 | 608 信号 | シゴウ | 信号 | 名詞一般 | | |

| index | filename | word | kana | part |
|-------|-----------------|------|------|------|
| 1 | 1 data001.txt | 信号 | シゴウ | 名詞一般 |
| 2 | 29 data001.txt | 信号 | シゴウ | 名詞一般 |
| 3 | 44 data001.txt | 信号 | シゴウ | 名詞一般 |
| 4 | 60 data001.txt | 信号 | シゴウ | 名詞一般 |
| 5 | 105 data001.txt | 信号 | シゴウ | 名詞一般 |
| 6 | 278 data002.txt | 信号 | シゴウ | 名詞一般 |
| 7 | 298 data002.txt | 信号 | シゴウ | 名詞一般 |
| 8 | 331 data003.txt | 信号 | シゴウ | 名詞一般 |
| 9 | 363 data003.txt | 信号 | シゴウ | 名詞一般 |
| 10 | 407 data003.txt | 信号 | シゴウ | 名詞一般 |
| 11 | 413 data003.txt | 信号 | シゴウ | 名詞一般 |

キーワード候補の表示では、count 数のみを表示し個々の情報については表示していません。詳細を表示したい場合カーソルを合わせて詳細ボタンを押してください。キーワード候補が出現する元文章の情報が表示されます。

10. 文章の表示



キーワード候補が出現してるファイルにカーソルを合わせて詳細ボタンを押してください。キーワード候補が出現してる元文章を表示します。