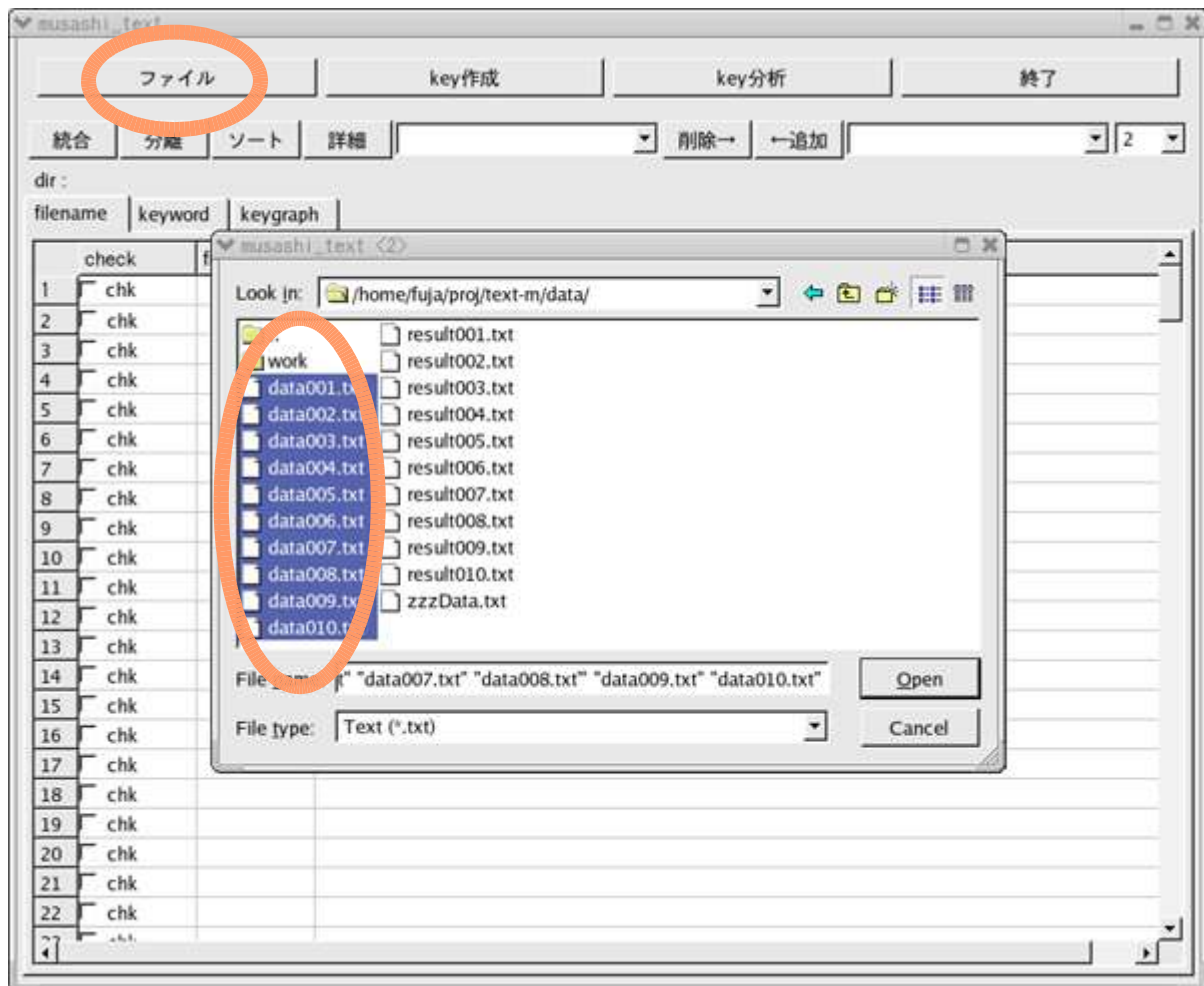
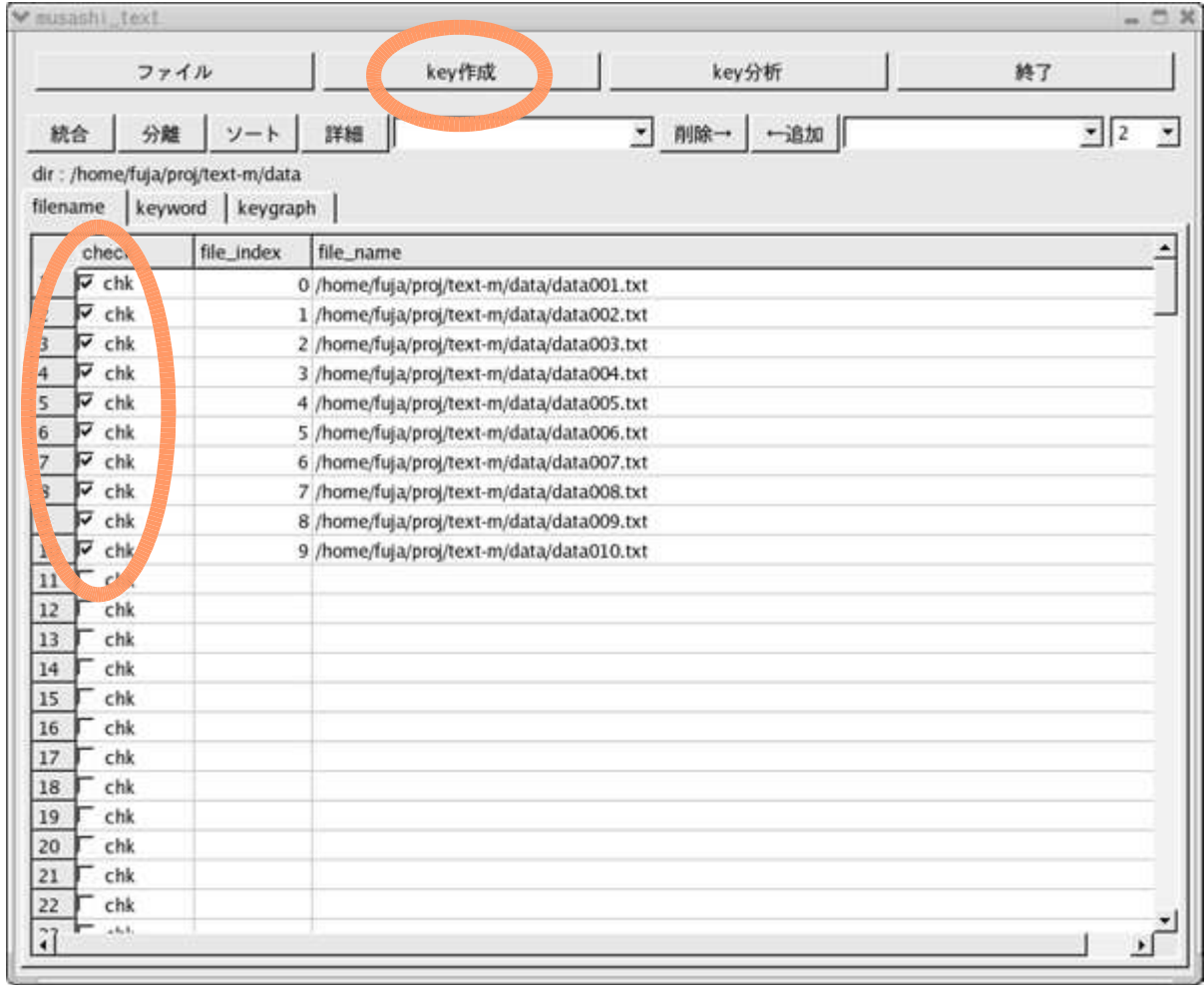


1. ファイルの指定



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 **euc-jp** を想定しています。必要であれば予め **euc-jp** へのコード変換をお願いします。

2. キーワード候補の作成



分析に用いるファイルには左のボックスにチェックを入れ(デフォルトでは入っていません)、key 作成ボタンを押してください。内部的に「茶筌」(形態素解析プログラム：<http://chasen.aist-nara.ac.jp/hiki/ChaSen/>)を利用したキーワード候補の作成が行われます。

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が、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



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



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す。

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

musashi_text

ファイル key作成 **key分析** 終了

統合 分離 ソート 詳細 名詞-サ変接続 削除 → ← 追加 副詞-一般 **4**

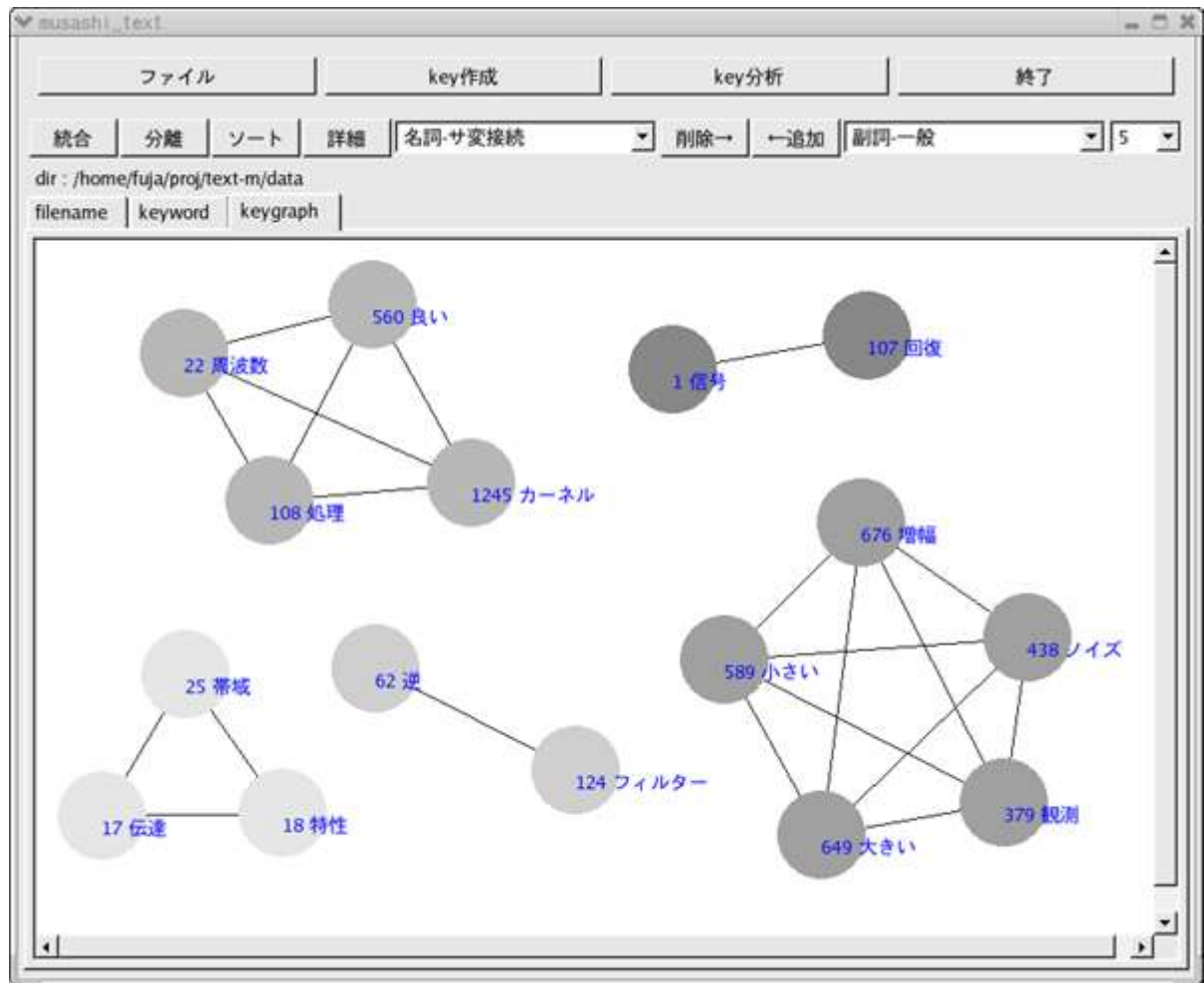
dir : /home/fuja/proj/text-m/data

filename keyword keygraph

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

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。

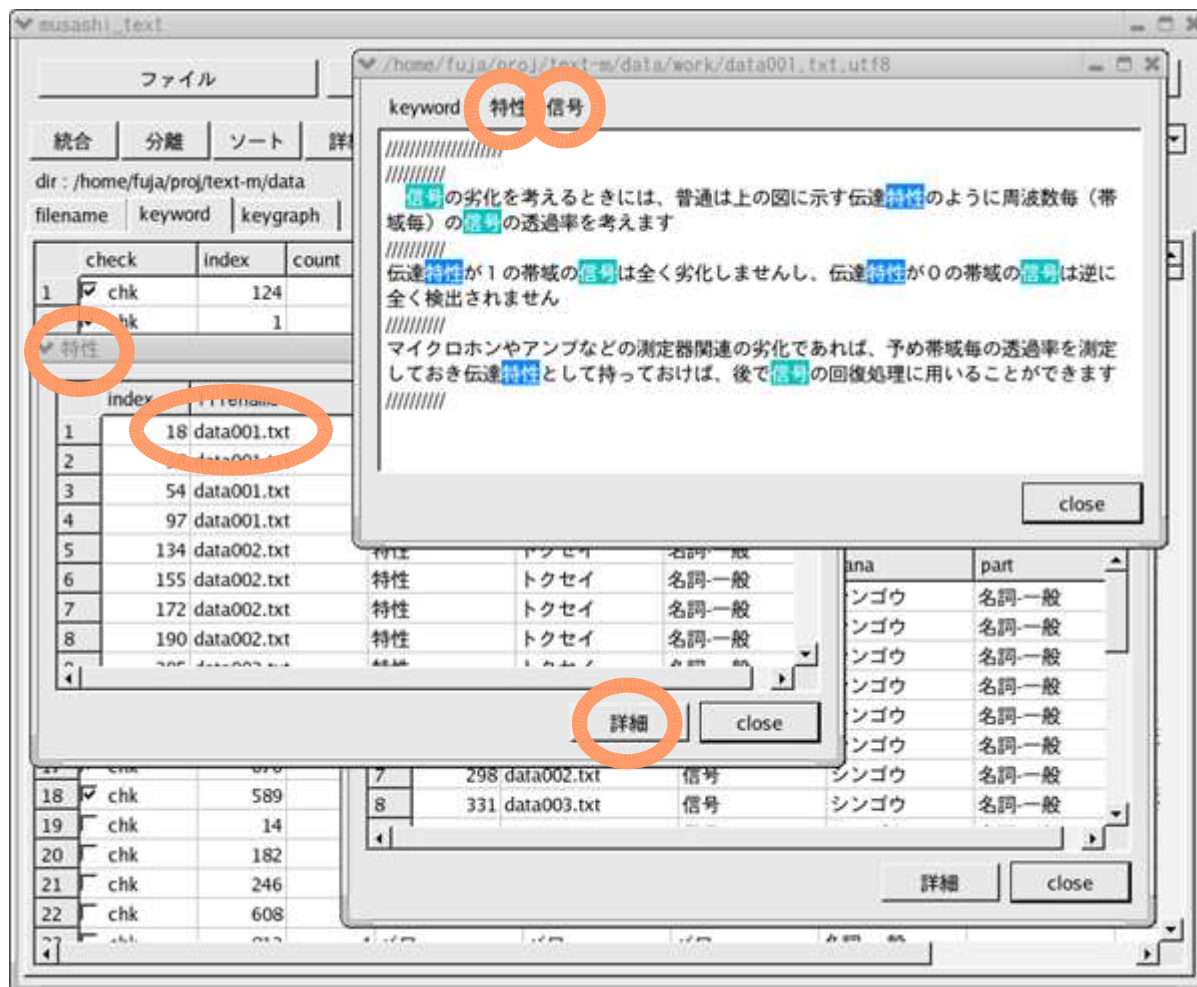
9. キーワードの詳細表示

The screenshot shows the 'musashi_text' application interface. The 'key作成' (Key Creation) tab is active. The '詳細' (Details) button is circled in red. A pop-up window titled 'musashi_text' is open, displaying a table of keyword details.

Index	filename	word	kana	part
1	1 data001.txt	信号	シグナル	名詞一般
2	29 data001.txt	信号	シグナル	名詞一般
3	44 data001.txt	信号	シグナル	名詞一般
4	60 data001.txt	信号	シグナル	名詞一般
5	105 data001.txt	信号	シグナル	名詞一般
6	278 data002.txt	信号	シグナル	名詞一般
7	298 data002.txt	信号	シグナル	名詞一般
8	331 data003.txt	信号	シグナル	名詞一般
9	363 data003.txt	信号	シグナル	名詞一般
10	407 data003.txt	信号	シグナル	名詞一般
11	413 data003.txt	信号	シグナル	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。