

Webページを活用した土木分野の情報技術利用調査

INVESTIGATION OF USE IN ENGINEERING WORKS FIELD OF INFORMATION TECHNOLOGY ON WEB PAGE

脇本文*, 河村圭**, 中村秀明***, 宮本文穂****

Fumi WAKIMOTO, Kei KAWAMURA, Hideaki NAKAMURA and Ayaho MIYAMOTO

*山口大学大学院 工学部 知能情報システム工学科 (〒755-8611 宇部市常盤台 2-16-1)

**工博 山口大学助手 工学部 知能情報システム工学科

***工博 山口大学助教授 工学部 知能情報システム工学科

****工博 山口大学教授 工学部 知能情報システム工学科

This paper deals with the Web Mining from JSCE Library's web page(<http://www.jsce.or.jp/library/>). JSCE Library's web page opens data of the thesis announced at academic meeting to the public with Web. We extract key words published in thesis from this web page. And we present a method for visualization of JSCE Library key words, as purpose investigation of relation between Engineering works technology and Information technology.

Key Words: Web Mining, JSCE Library, Web graph, Information technology

1. はじめに

インターネットは、膨大なデータを有し、誰でも利用可能なデータベースとみなすことができる。このデータを有効に活用することは大きな利益に繋がることから、インターネットより情報の抽出を行うWebマイニングが注目を集めている¹⁾。一方、土木分野においてGA、ニューラルネットワークなどの情報技術の普及が進んでいる。また、各自治体、公団においてデータベースの整備などが進み、電子情報の蓄積が行われていることから、これらの技術は今後もより広く利用されていくと考えられ、利用状況の現状を把握しておくことは重要である。

また、土木学会附属土木図書館では、学会において発表された論文の電子データ化が行われており、このデータはWeb上で公開されている。そこで本研究では、土木学会附属土木図書館Webページより土木学会において発表された論文から技術用語を取得する。取得した用語間の関係を調査することで、土木において情報技術がどのように利用されているのか、調査を行う。

以前にも、AL技術の土木工学における広がり²⁾と題して、土木学会附属土木図書館よりデータを収集し、情報技術、特にAL（人工生命）技術の土木分野への適用調査がおこなわれているが、それはAL技術を使用した論文数の調査、および土木学会における各部門（I～VII）における個別の集計にとどまるものであった。今回はより幅広く詳しい調査を目的とし、土木分野に適用されている情報技術は何か、また具体的にどの領域に適用されているかを明らかにすることを目的とする。

2. キーワードの取得

土木学会附属土木図書館Webページの「目録・書誌検索」を利用し、土木学会誌、土木学会論文集、土木学会委員会論文集（全国大会年講を含む）および支部論文集より検索して文献を探した。土木学会附属土木図書館Webページには、論文の抄録データが図1のようにHTMLで保存されており、検索語句がHTML上のどこかに含まれている論文が検索にヒットする。今回は、情報技術の使用された論文を幅広く検索するために、検索語句として『情報』を使用した。こうして得られた文献のページからHTMLタブを参照することにより、「キーワード（和）」の行から技術用語を取得し、調査に使用した。

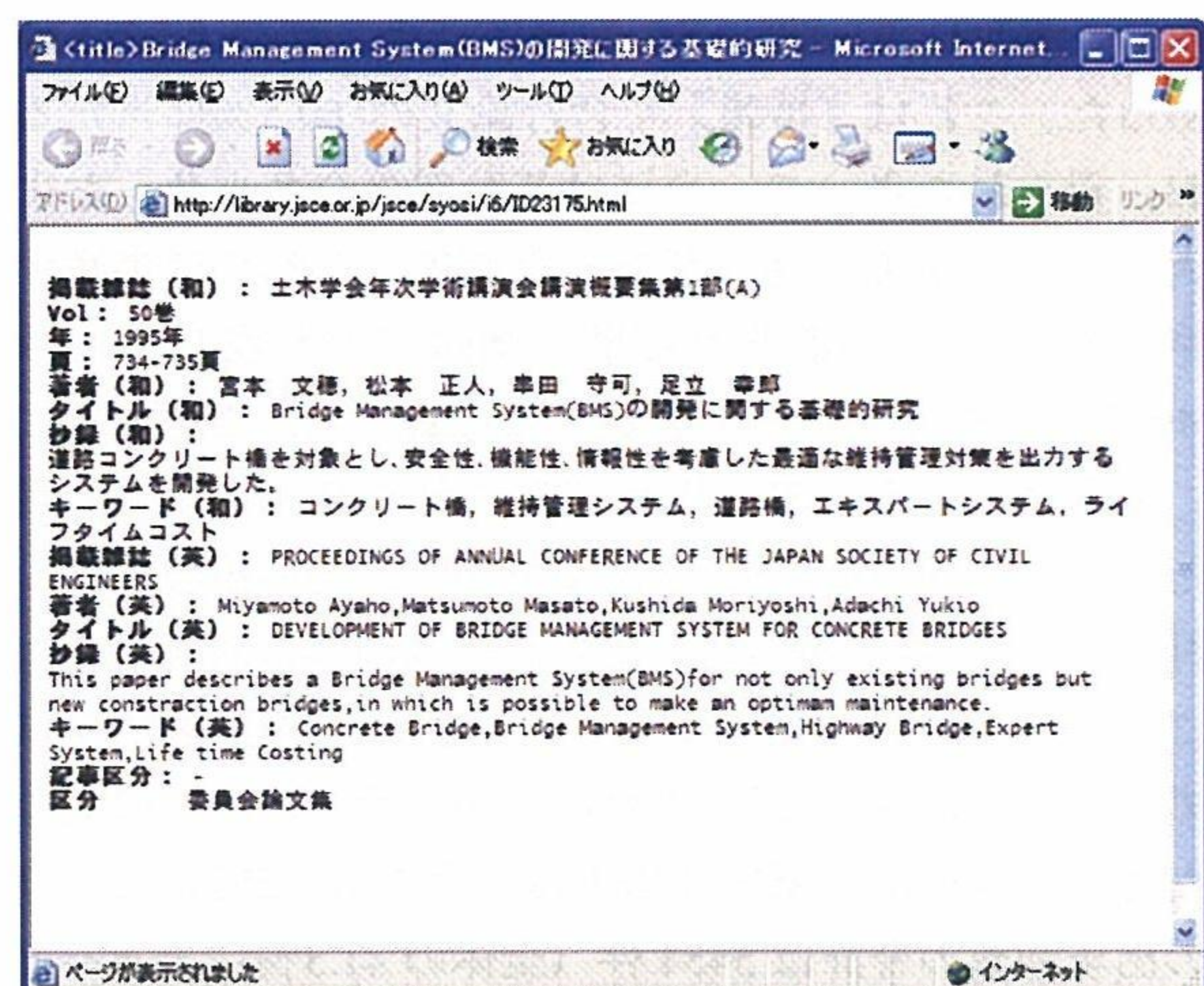


図1 土木学会附属土木図書館、論文の抄録ページ

表1 キーワードの出現頻度(件数4件以上)

件数	キーワード
198	交通情報
155	GIS
144	地理情報システム
142	データベース
122	情報化施工
107	計画情報
96	ITS
75	インターネット
65	シミュレーション
65	GPS
65	逆解析
61	交通行動分析
56	ニューラルネットワーク
56	CAD
55	情報提供
55	経路選択
48	リモートセンシング
48	防災計画
46	維持管理
43	交通安全
43	地理情報
42	CALS
41	交通管理
41	施工管理
40	情報処理

表2 23 カテゴリ

岩盤・地盤	GIS	情報
治水	橋梁	情報機器
設計	住民・合意形成	情報通信
施工	交通	CAD
維持管理	計測モニタリング	CALS
景観	都市計画	標準化
環境	情報手法	マネジメント
防災	情報理論	

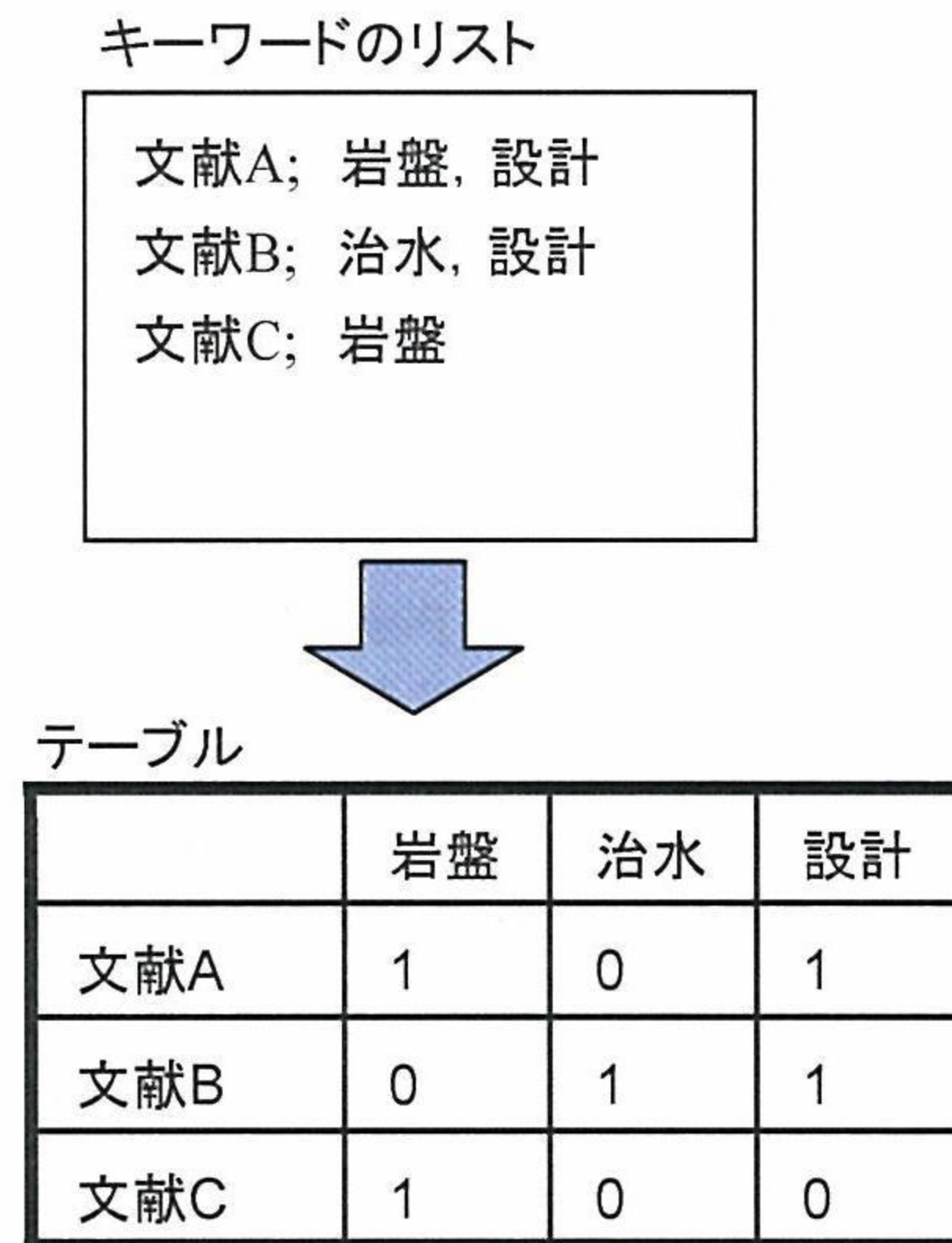


図2 Webグラフ用レコードの作成

検索対象となる論文の総数は約 23 万件である。検索語句『情報』でヒットした論文の件数は 5641 件であり（平成 17 年 7 月現在），総数からするとごく少数であった。本ページの検索システムの仕様の都合上，この検索結果では十分に情報技術を含む論文を網羅できていないことが理由のひとつとして挙げられる。別の検索方法を検討中である。

上記の手段を通して得られた文献の中には，取得する「キーワード（和）」が記述されていないものが存在しており，実際にキーワードの取得ができた論文は 3415 件であった。これは『情報』を含む論文の 60.5% に当たる。なお，得られたキーワードには英数字の全角半角，および大文字小文字での違いがあったため，これらは半角大文字に統一した。集計の結果，全体での出現した件数が 40 件以上のキーワードを表 1 に記載した。これ以外にも，取得されたキーワードは多岐にわたり，全体で取得したキーワードの種類は 6202 種であり，その出現件数の合計は 13200 件であった。出現が一度のみの用語などは特に多く，集計が困難であったため，今後の解析は全体での出現回数が 4 件以上の物について行った。その対象となるキーワードの種類は 464 種（全体の 7.5%）であり，その出現回数の合計は 5454 件（全体の 41.3%）であった。

3. Web グラフの作成

情報技術が応用されているのはどの分野であるかを確認するために，SPSS クレメンタインの Web グラフ³⁾によって視覚化を行う。これは，共起関係（頻繁に特定の単語と単語が同時に出現すること）のある単語同士を線で結び，その関係の強さを線の太さによって表すグラフである。このグラフを使用することで，複数の論文のキーワード欄において，同時に出現しているキーワードを発見することが可能になる。それにより，それぞれのキーワードが表す研究の関係を視覚化し，どのような研究分野が，どのような研究分野に適用されているか調査できる。

具体的に Web グラフを作成するには，図 2 のようなテーブルを作成することになる。これは多変量解析におけるダミー変数のようなもので，文献のキーワードに『岩盤』が含まれているのならば，『岩盤』の項目に 1 の値が入り，そうでなければ 0 の値が入る。このテーブルの項目はキーワードの種類の数だけ作成される。しかし，対象としているキーワード全 464 種をそのまま用いたのでは，キーワードの種類が膨大な事からグラフが複雑になり，関係を読み取ることが著しく困難になる。そこで，キーワードを大まかに分類してからグラフ用のデータを作成した。まず，各キーワードを意味，分野等

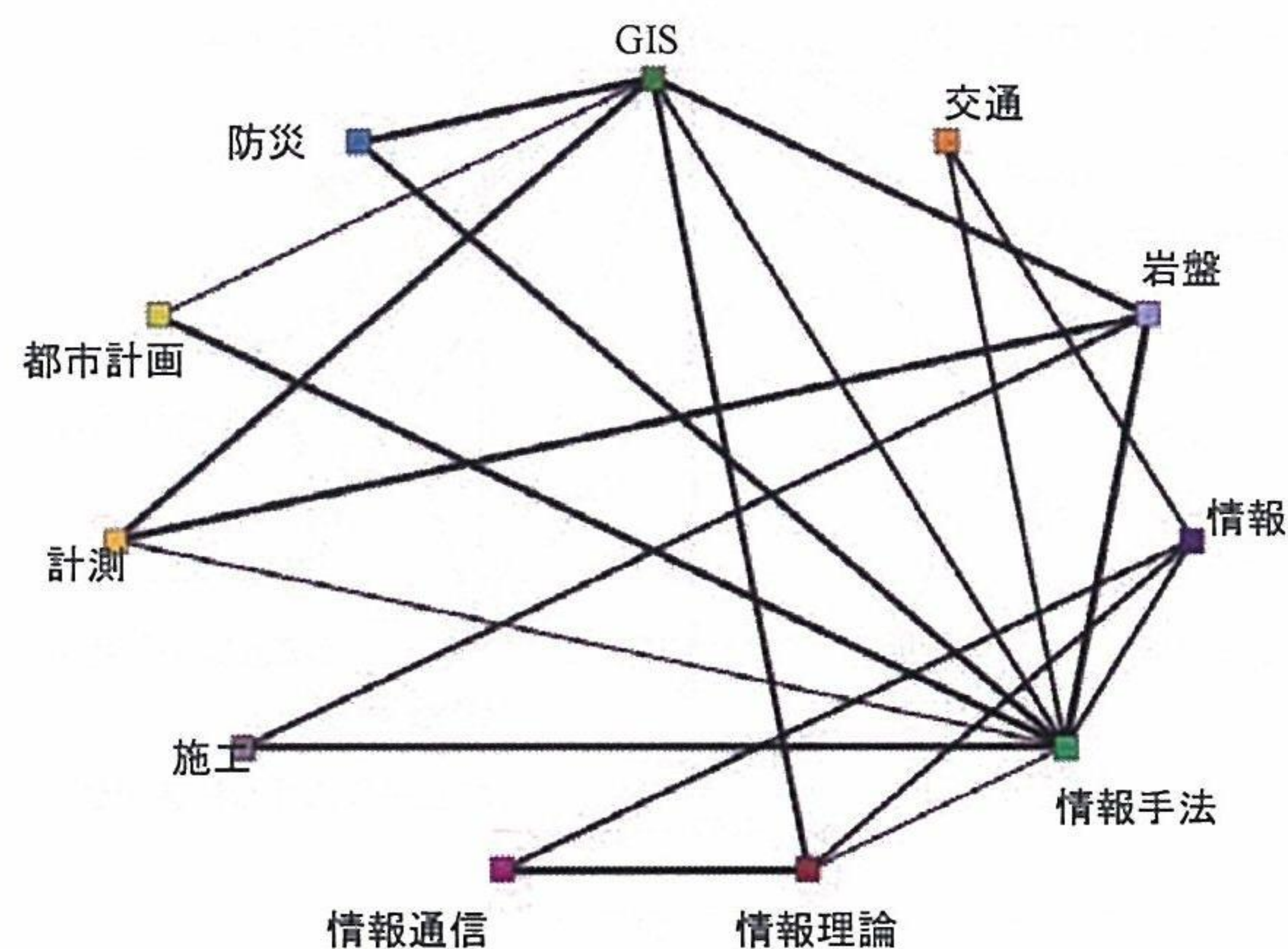


図3 Web グラフ

の近さでいくつかのグループに振り分けていき、決定したグループに含まれるキーワードから大分類の呼称を決定した。最終的に作成した大分類は23種となった。その内訳は表2のとおりである。なお、分類が仕切れず、どの大分類にも含まれなかったキーワードがあることを断っておく。

上記の分類にしたがって作成したWebグラフが図3となる。すべての大分類間の関係を一度に示せばグラフが煩雑になり、関係を読み取ることが不可能になるため、図3ではグラフの見易さを優先し、特に強い関係のみ上位20件を表示した。

図3において、情報関連の技術「GIS」、 「情報通信」、 「情報理論」、 「情報手法」および「情報」の節点と線で結ばれている節点は、「交通」、 「岩盤・地盤」、 「施工」、 「計測モニタリング」、 「都市計画」および「防災」である。このことから、土木学会において情報技術が用いられている分野は「交通」、 「岩盤・地盤」、 「施工」、 「計測モニタリング」、 「都市計画」および「防災」であるといえる。

ここで、Webグラフはあくまで一対一の関係を示すグラフであることを注意しておく必要がある。例えば、図3において「GIS」、 「防災」また「情報手法」が辺の太い三角形の各頂点をなしているが、これは「GIS」と「防災」、 「防災」と「情報手法」、 また「情報手法」と「GIS」の一対一の関係が深いことを示すのみで、必ずしも三つが同時に関係しているわけではない。

4. 各大分類間の細かい関係

前節で示したWebグラフにより、大分類間の関係を視覚的に表示することができた。しかし、前節では大分類の意味するところが広く、大分類間に関係があるとわかっても、それが具体的にどのような関係なのか不明である。

そこで大分類間の関係をより深く調べるために、前節において特に強い関係があるとされた二つの大分類につ

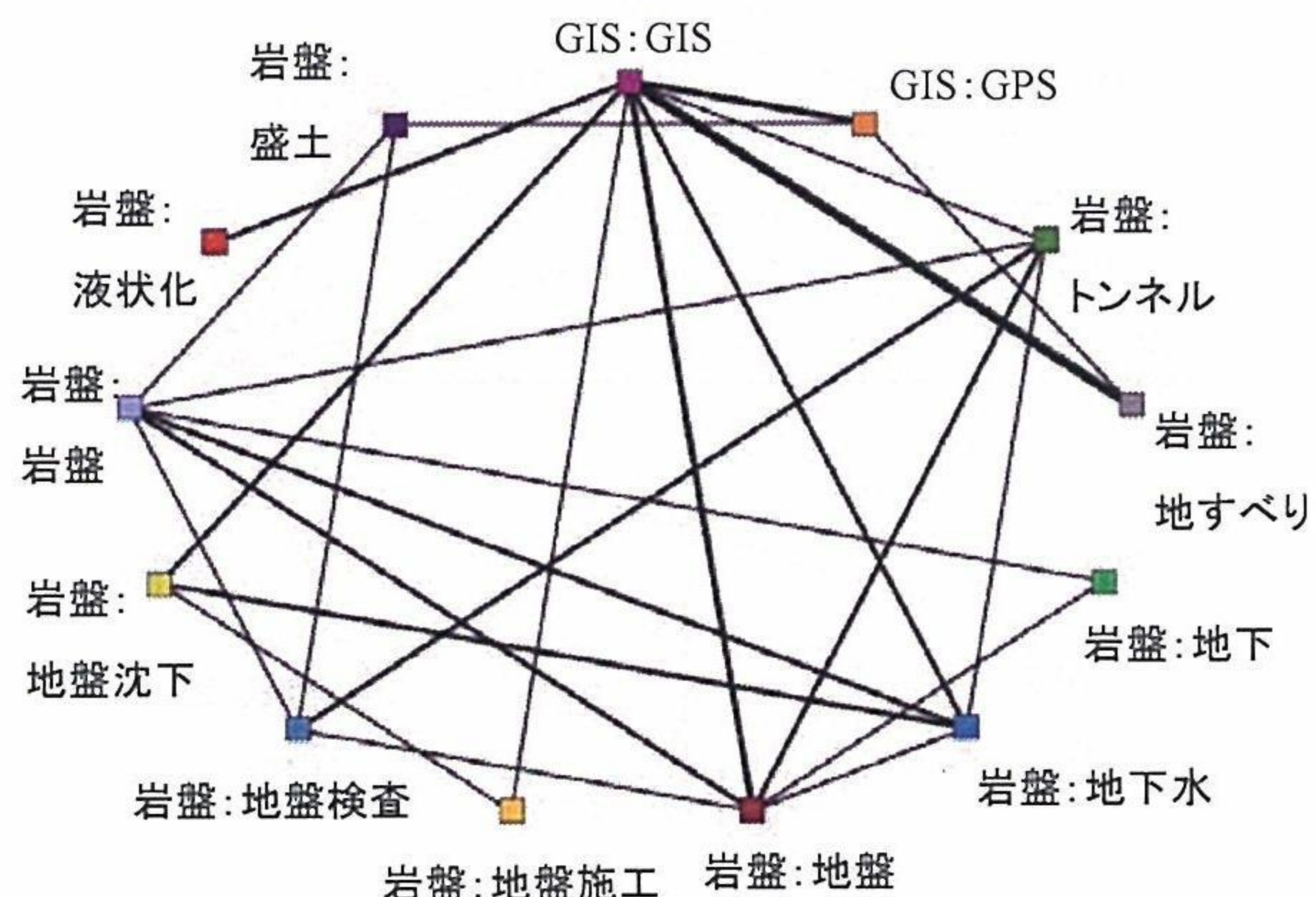


図4-1 GISと岩盤・地盤の詳細なWebグラフ

いてのWebグラフを作成する。まず、関係を調べたい大分類に含まれるキーワードを細かく再分類し、その小分類についてのWebグラフを作成する。

まず、キーワード単体としても出現頻度が高く、図3においても多くの大分類と強い関係を示した「GIS」に着目し、「GIS」と強い関係を示した「岩盤・地盤」、 「情報理論」、 「計測モニタリング」、 「防災」および「情報手法」との関係調べた。

図4-1から図4-5に、それぞれ「GIS」と「岩盤・地盤」、 「GIS」と「情報理論」、 「GIS」と「計測モニタリング」、 「GIS」と「防災」および「GIS」と「情報手法」のキーワードの小分類に従ったWebグラフの結果を示した。表示する関係は2件以上のものを示している。図中のノードは「岩盤:盛土」のように表記されているが、これは大分類では「岩盤」に分類されていたもので、今回の小分類で「盛土」に分類されたものを指している。また、都合により「岩盤・地盤」→「岩盤」、 「計測モニタリング」→「計測」のように、大分類名を省略しているところもある。

図4-1から、「GIS」が「岩盤・地盤」で主に関係しているのは斜面の安定性や崩壊の危険度などの研究である「地すべり」であることがわかる。

図4-2において「GIS」と「DB」の関係が強く現れているのは、地理情報をいかに保存するかに関心が高いことの表れである。

図4-3から、「リモートセンシング」の技術と「GIS」に強い関係があることがわかる。「リモートセンシング」と、その次に「GIS」と強い関係を示す「空中衛星写真」にはともに人工衛星に関連する技術が多いことから、人工衛星と「GIS」に深い関係があると捉えることができる。

図4-4より、「地震」が「GIS」と強い関係を示していることがわかる。「地震」は防災分野において特に関心もたれており、キーワードとして取得された件数が顕著であった分野なので、「GIS」と「防災」との関係

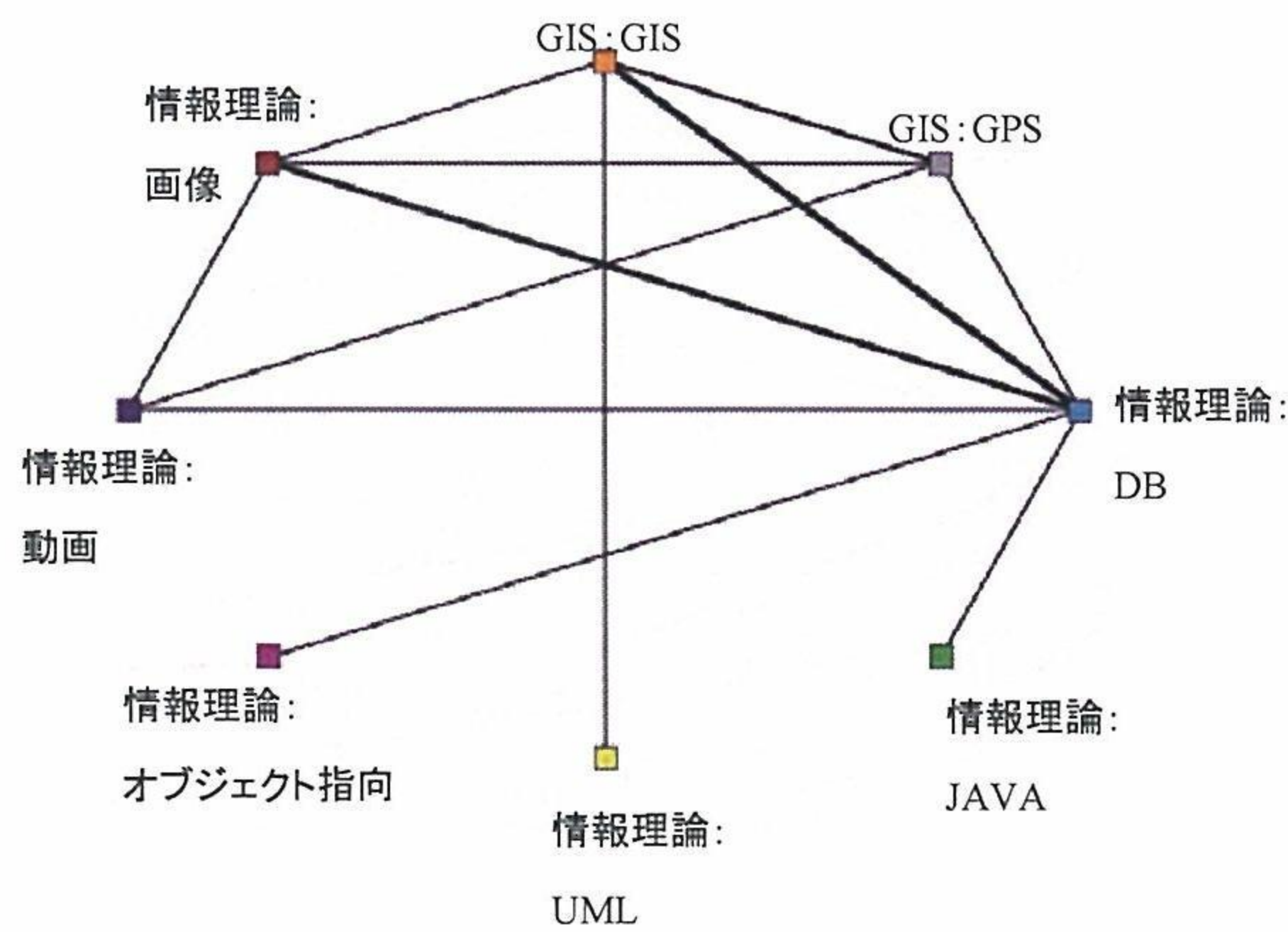


図 4-2 GIS と情報理論の詳細な Web グラフ

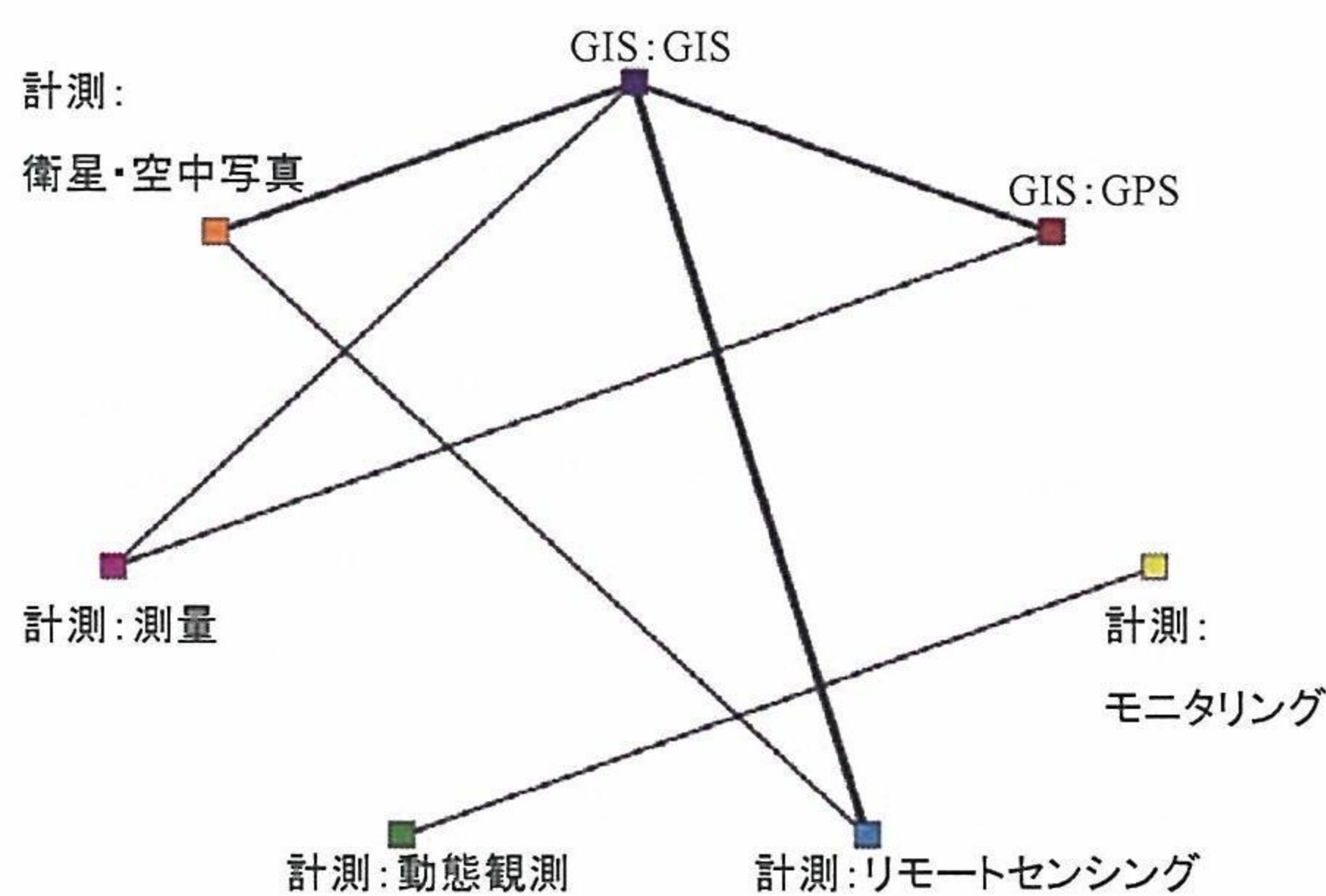


図 4-3 GIS と計測モニタリングの詳細な Web グラフ

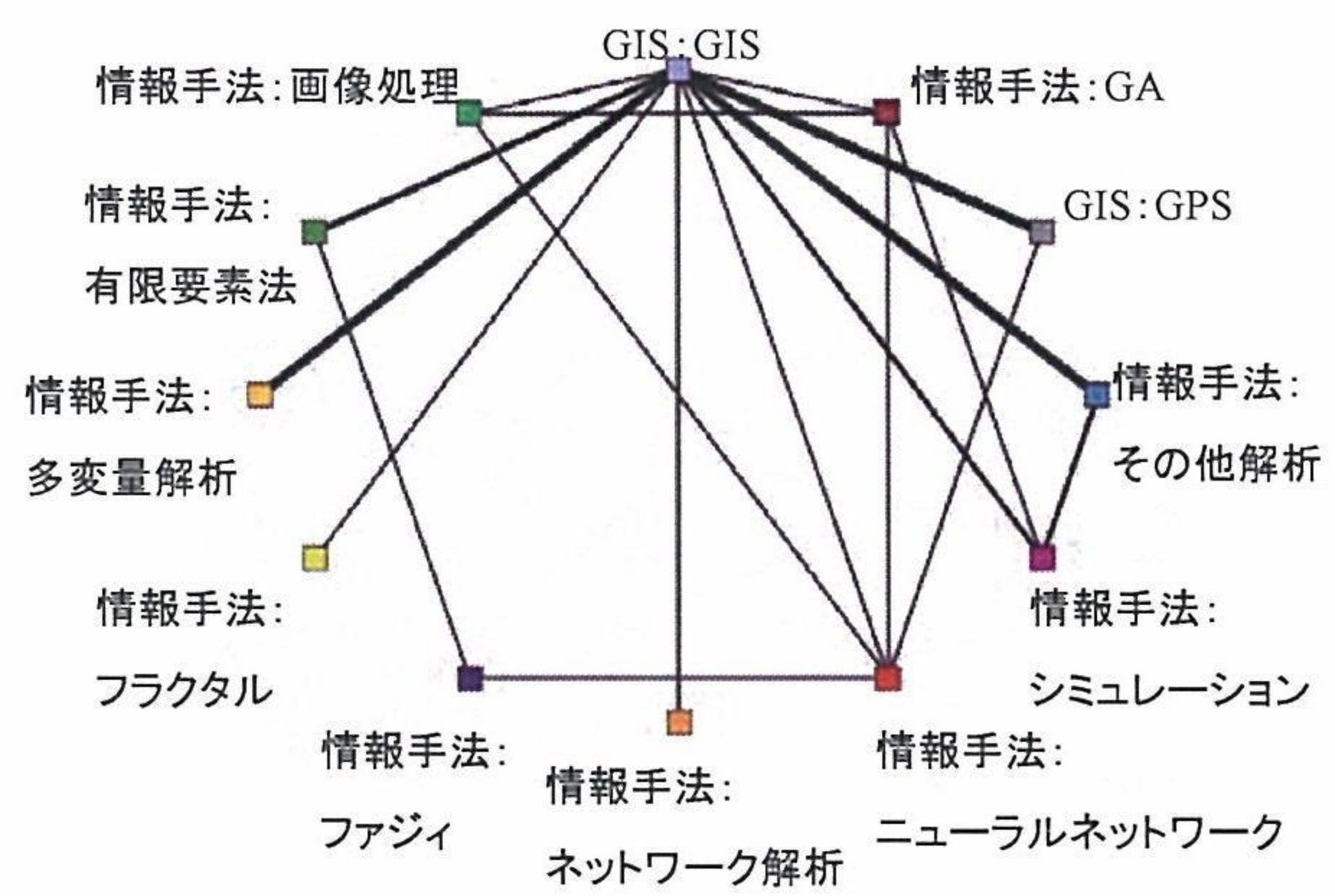


図 4-4 GIS と防災の詳細な Web グラフ

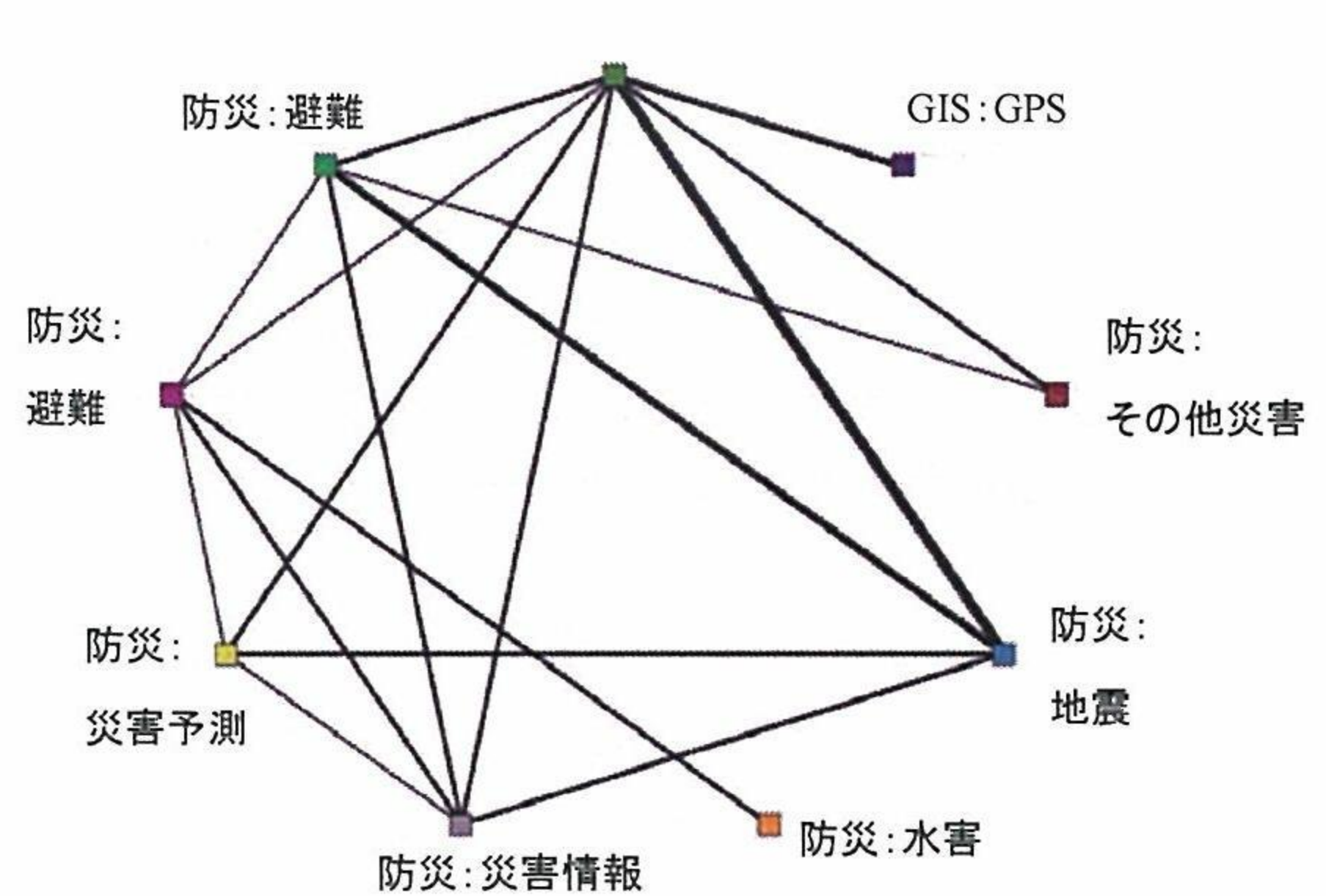


図 4-5 GIS と情報手法の詳細な Web グラフ

の大半を占めているのも頷ける。しかし、「地震」についてどのように「GIS」が使われているのかがこの図からは不明なので、さらに詳しく調べる必要がある。

図 4-5 より、「GIS」と「情報手法」の内訳は程よく分散されていることがわかる。しかしその中でも、「有限要素法」が使われることが特に多いようである。GIS と関係を持った「その他解析」の内訳は「計画情報」がほぼ 7 割を占めており、残りはほぼ「階層分析法」であった。

5. 結論

本研究では、土木学会附属土木図書館 Web ページより、論文のデータを取得し、そのキーワード間の関係を調べることで、土木分野における情報技術の適用状況の調査を行った。キーワードをいくつかの分類に分け、Web グラフを使うことにより、大まかな適用状況の把握を行った。それによって関係があるとされた分類内の細かい関係を調べることにより、さらに詳しい適用状況の把握

を行った。以下、今後の展望について述べる。

今回は論文中のキーワードのみを対象として調査を行ったが、さらに詳しい調査を行うため、論文のタイトル、および抄録の自由文を解析することが考えられる。論文の発表年から、年代ごとに情報を分け、情報技術適用の遍歴を調査することも考えられる。また、より広い範囲での調査のため、土木学会附属土木図書館以外からの情報の取得が考えられる。

参考文献

- 1) 坂本比呂志, 有村博紀, Web マイニング, 特集「テキストマイニング」, 人工知能学会誌, Vol.16, No.2, 2001年3月..
- 2) 土木学会 構造工学委員会 AL(人工生命)技術の構造システム最適化への応用に関する研究小委員会: 活動成果報告書 第4期 1999.4-2002.3, 2001.10
- 3) SPSS 社: Clementine 8.0 User's Guide,