

数字で見る全国遺跡報告総覧

高田祐一（奈良文化財研究所）

The Progress of the Comprehensive Database of Archaeological Site Reports in Japan

Takata Yuichi (Nara National Research Institute for Cultural Properties)

・データ量／Amount of data ・自然言語処理／NLP

全国遺跡報告総覧の経過とデータ件数などについて紹介する。

1. 現在のデータ件数

全国遺跡報告総覧に登録されているデータ件数は以下の通り（2019年12月25日時点）。推移は表1。

PDFがある書誌登録数：23,970

PDF総ページ数：2,984,610

PDF総文字数：1,871,574,794

書誌登録数：63,988

書誌の発行機関数：1,596

遺跡抄録件数：128,072

2. 全国遺跡報告総覧のこれまでの経過

2008年から2014年は遺跡資料リポジトリである。

2008（平成20）年度

- ・中国地方5県域（大学）で遺跡資料リポジトリの運用開始（各大学が個別にサーバ管理）
- ・メタデータの収集

・作成開始（奈文研とも連携）

・報告書の電子化開始（NIIのCSI事業費）

2009（平成21）年度

- ・12府県域（大学）へ拡大（各大学が個別にサーバ管理）
- ・横断検索・全文検索に対応するためのシステム改修
- ・報告書の電子化（CSI事業費／科研費）

2010（平成22）年度

- ・20府県域（大学）へ拡大
- ・NII共同サーバ上にシステム構築（新規参加大学向け）
- ・オープンカンファレンス開催（2010.12@大阪）
- ・報告書の電子化（CSI事業費／科研費）

2011（平成23）年度

- ・自治体セルフ・アーカイブ移行のためのシステム改修
- ・全国の自治体にアンケート調査を実施
- ・プロジェクトが推進する電子化仕様を公開
- ・ワークショップ開催（2010.11@東京／同.12@福岡）
- ・報告書の電子化（CSI事業費）

2012（平成24）年度

- ・CSI委託事業最終年度
- ・22府県域（21大学）へ拡大

表1 全国遺跡報告総覧に登録されているデータ件数の推移（2019年12月15日時点）

| 年度 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 (途中) |
|---------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|--------------|
| PDF登録件数 | 920 | 4552 | 7525 | 9771 | 11916 | 13516 | 14374 | 15930 | 18838 | 21154 | 23075 | 23970 |
| 書誌登録数 | 920 | 4552 | 7525 | 9771 | 11916 | 13516 | 14374 | 15930 | 18838 | 21154 | 48694 | 63988 |
| 抄録登録数 | | | | | | | | | | 35000 | 93923 | 128077 |
| 直接登録機関数 | | | | | 6 | 22 | 23 | 31 | 50 | 875 | 910 | 1235 |

※2019年度は2019年12月25日時点。2008年度から2014年度は遺跡資料リポジトリの統計。2017年度抄録登録数は概算

- ・報告書の電子化（CSI事業費／科研費）
- ・奈文研と広域モデル実証実験開始（報告書発行機関による直接参加が可能に）
- ・シンポジウム開催（2012.11@福岡）

2013（平成25）年度

- ・報告書の電子化（科研費）
- ・奈文研とシステム移行に向けた協議開始

2014（平成26）年度

- ・島根大学附属図書館全国遺跡資料リポジトリ推進会議事務局が国立大学図書館協会賞を受賞（2014.6）
- ・連携大学実務者会議でシステムの統合・移管が決定（2014.11）
- ・旧システムの凍結及びシステム移行作業（2014.12～）

2015（平成27）年度

- ・連携大学による新システムの検証（負荷テスト等）
- ・全国遺跡報告総覧の公開（公開時点の報告書14,374冊）（2015.6）
- ・データ登録再開（2015.8）
- ・ディスカバリーサービス（Summon）との連携開始（2015.9）
- ・報告書本文データの登録件数が15,000件に（2015.10）
- ・シンポジウム開催（2016.2@奈良）
- ・CiNii Booksとの連携開始（2016.3）

2016（平成28）年度

- ・英語自動検索機能公開（2016.8）
- ・イベント情報（文化財イベントナビ）の登録・公開機能を追加（2016.9）
- ・シンポジウム開催（2016.11@奈良）
- ・Worldcat（ディスカバリーサービス含む）との連携開始（2017.2）

2017（平成29）年度

- ・ディスカバリーサービス（EDS）との連携開始（2017.4）
- ・報告書の頻出用語を可視化したワードマップを公開（2017.4）
- ・報告書本文データの登録件数が20,000件に

（2017.7）

- ・文化財報告書にDOIの付与開始（2017.7）
- ・考古学関係用語辞書拡充（2017.8）
- ・「データ登録に関する今後の方針」を公開（2017.10）
- ・報告書発行機関向けの説明会を開催（5会場：奈良／仙台／岡山／福岡／東京）

2018（平成30）年度

- ・モバイル端末向けPDFの公開（2018.8）
- ・遺跡（抄録）検索機能の公開（2018.12）
- ・ディスカバリーサービス（Primo）との連携開始（2018.12）
- ・全埋協抄録データベースの統合完了（2019.1）
- ・引用表記の自動表示（2019.2）
- ・都道府県別の発掘調査報告書総目録 高知県・島根県編の公開（2019.3）
- ・デジタル技術導入に関する研究報告を全国遺跡報告総覧で公開（2019.3）
- ・報告書発行機関向けの説明会を開催（5会場：京都／福島／石川／埼玉／大分）

2019（令和元）年度

- ・欧州考古学情報基盤 ARIADNE Plusへの奈良文化財研究所の参画（2019.4）
- ・都道府県別の発掘調査報告書総目録 新潟県編の公開（2019.4）
- ・全国の遺跡や文化財に関するイベント情報検索機能公開（2019.6）
- ・都道府県別の発掘調査報告書総目録 大阪府編の公開（2019.6）
- ・奈文研抄録データベースの全国遺跡報告総覧への統合完了（2019.6）
- ・発掘調査報告書総目録 新潟県編の書誌情報を全国遺跡報告総覧に登録（2019.11）
- ・書誌ページQRコード表示機能とシリーズ番号順並び替え機能公開（2019.11）
- ・報告書発行機関向けの説明会を開催（5会場：佐賀／新潟／広島／愛知／東京）

3. 利用統計

2018年度の1年間のPDFダウンロード数は1416171件であった(図1)。アクセス数は1117万件、ページ閲覧数は1億302万件だった(表2)。

4. 自然言語処理からみた報告書データ

4.1 頻出用語俯瞰図

全国遺跡報告総覧には日本語18億文字が登録されている。自然言語処理にて図化した結果を示す。

図2の報告書ワードマップ(頻出用語俯瞰図)は全国遺跡報告総覧に登録されている報告書に対し、考古学関係用語の出現回数を集計し、図化した。用語については桃色：遺物に関する用語、黄色：遺構に関する用語、水色：その他で分類した。最新の図は全国遺跡報告総覧「報告書ワードマップ」(<https://sitereports.nabunken.go.jp/ja/visualization/term>)で閲覧できる。



図1 報告書登録数とダウンロード数の推移

表2 全国遺跡報告総覧のアクセス数とページ閲覧数

| 年度 | アクセス数 | ページ閲覧 |
|------|-------|--------|
| 2015 | - | - |
| 2016 | 341万 | 1155万 |
| 2017 | 886万 | 7277万 |
| 2018 | 1117万 | 1億302万 |

4.2 都道府県別特徴語

全国遺跡報告総覧に登録されている報告書に対し、都道府県ごとに考古学関係用語の特徴語を図化した。当該都道府県内にて頻出する用語(よく使われる用語は重要)かつ他都道府県では出現頻度が低い用語(希少用語は重要)であることを勘案するため、当該都道府県の強い特徴を示す用語を可視化できる。自然言語処理技術のベクトル空間モデルのTF(索引語頻度)とIDF(逆文書頻度)を組み合わせたTF-IDFにて算出した(図3～26)。データ数が少ない場合、結果が偏る可能性があるため、PDFが登録されている報告書類(全国遺跡報告総覧の報告書種別a01)について、都道府県下で300件を超えた場合にワードマップを作成することとなっている。最新の図は「報告書ワードマップ」(<https://sitereports.nabunken.go.jp/ja/visualization/term>)にて都道府県を選択すると閲覧できる。

5. 都道府県の状況

都道府県別のPDF・書誌・抄録公開件数を表3に示す。2019年12月25日時点。最新の情報は「発行機関一覧(都道府県別)」(<https://sitereports.nabunken.go.jp/ja/list>)で閲覧できる。

表3 全国遺跡報告総覧の都道府県別PDF・書誌・抄録公開件数

| | PDF公開 | 書誌のみ公開 | 書誌合計 | 抄録 |
|------|-------|--------|-------|--------|
| 北海道 | 95 | 1774 | 1869 | 2750 |
| 青森県 | 212 | 1284 | 1496 | 3419 |
| 岩手県 | 585 | 1332 | 1917 | 3280 |
| 宮城県 | 1147 | 376 | 1523 | 1097 |
| 秋田県 | 601 | 198 | 799 | 2320 |
| 山形県 | 787 | 188 | 975 | 1299 |
| 福島県 | 0 | 1616 | 1616 | 4066 |
| 茨城県 | 749 | 694 | 1443 | 2515 |
| 栃木県 | 50 | 565 | 615 | 1078 |
| 群馬県 | 628 | 2205 | 2833 | 5011 |
| 埼玉県 | 241 | 2230 | 2471 | 4950 |
| 千葉県 | 85 | 3397 | 3482 | 7815 |
| 東京都 | 16 | 1870 | 1886 | 3836 |
| 神奈川県 | 87 | 850 | 937 | 1579 |
| 新潟県 | 372 | 2064 | 2436 | 3034 |
| 富山県 | 852 | 173 | 1025 | 1874 |
| 石川県 | 81 | 824 | 905 | 1248 |
| 福井県 | 4 | 298 | 302 | 435 |
| 山梨県 | 1083 | 84 | 1167 | 705 |
| 長野県 | 2680 | 415 | 3095 | 3925 |
| 岐阜県 | 204 | 466 | 670 | 1852 |
| 静岡県 | 452 | 1014 | 1466 | 2935 |
| 愛知県 | 250 | 1000 | 1250 | 2162 |
| 三重県 | 657 | 615 | 1272 | 2397 |
| 滋賀県 | 262 | 714 | 976 | 2254 |
| 京都府 | 151 | 1252 | 1403 | 4894 |
| 大阪府 | 1795 | 1830 | 3625 | 12535 |
| 兵庫県 | 724 | 887 | 1611 | 4202 |
| 奈良県 | 1173 | 852 | 2025 | 3811 |
| 和歌山県 | 91 | 332 | 423 | 1600 |
| 鳥取県 | 329 | 512 | 841 | 1834 |
| 島根県 | 1181 | 194 | 1375 | 3655 |
| 岡山県 | 632 | 124 | 756 | 990 |
| 広島県 | 87 | 668 | 755 | 1532 |
| 山口県 | 39 | 570 | 609 | 963 |
| 徳島県 | 121 | 104 | 225 | 405 |
| 香川県 | 634 | 176 | 810 | 1899 |
| 愛媛県 | 222 | 493 | 715 | 1180 |
| 高知県 | 388 | 42 | 430 | 474 |
| 福岡県 | 1400 | 2824 | 4224 | 6988 |
| 佐賀県 | 153 | 419 | 572 | 1187 |
| 長崎県 | 356 | 294 | 650 | 1040 |
| 熊本県 | 380 | 285 | 665 | 1206 |
| 大分県 | 54 | 777 | 831 | 1761 |
| 宮崎県 | 1196 | 124 | 1320 | 2604 |
| 鹿児島県 | 466 | 610 | 1076 | 1724 |
| 沖縄県 | 218 | 403 | 621 | 838 |
| 合計 | 23970 | 40018 | 63988 | 125158 |

※遺跡抄録にて市町村コードがないものはカウント外のため、全体の件数と一致しない

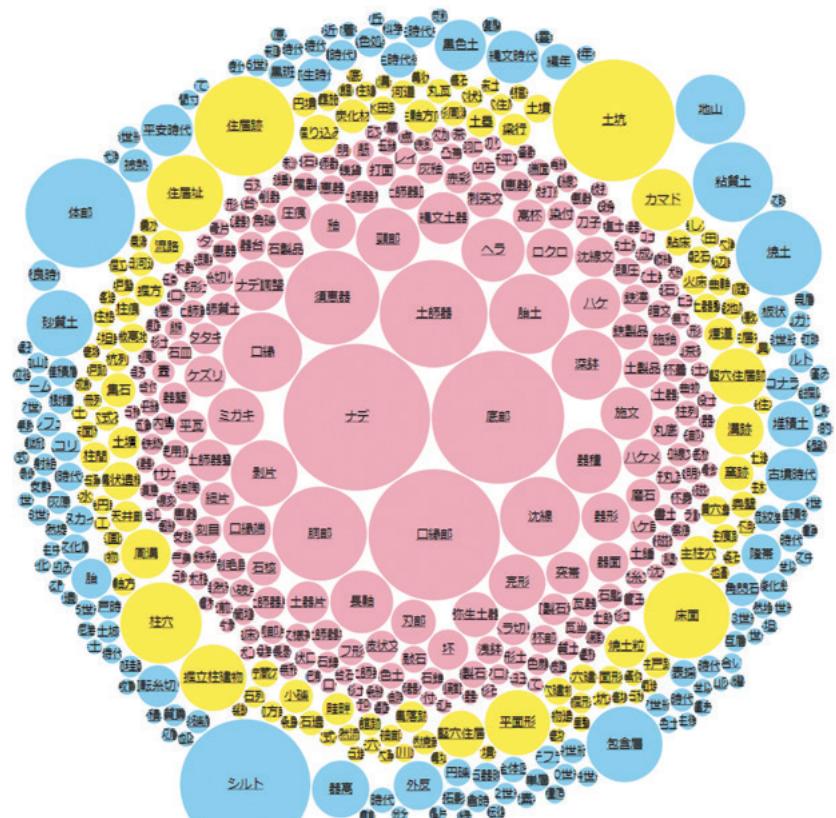


図2 報告書ワードマップ（頻出用語俯瞰図）
 (https://sitereports.nabunken.go.jp/ja/visualization/term) 2019年12月25日時点。

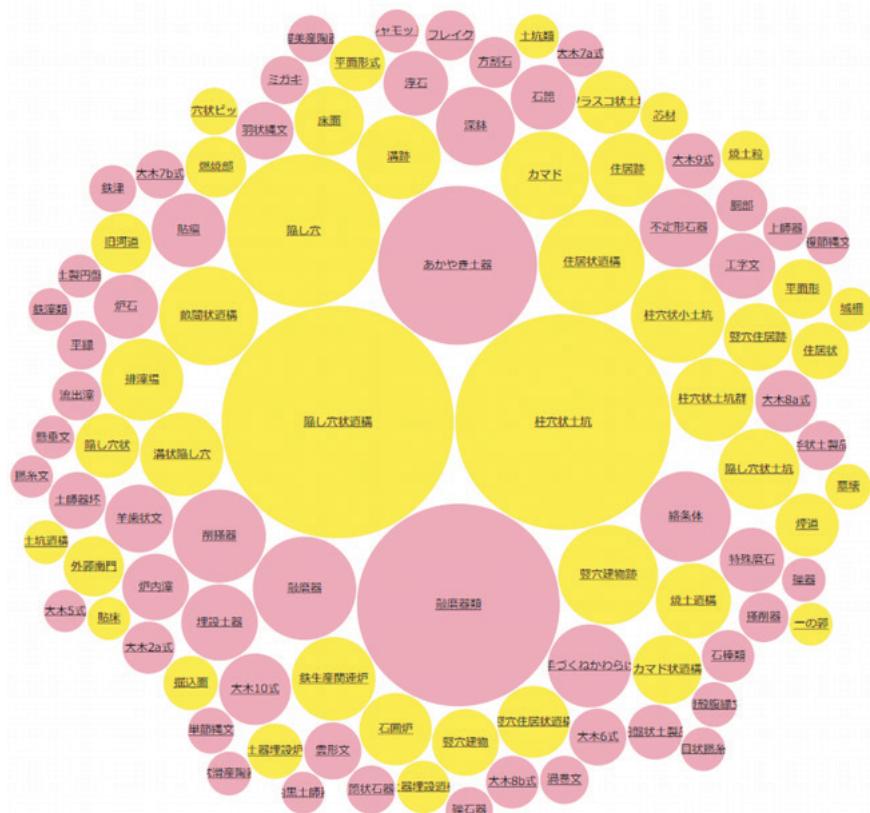


図3 報告書特徴語ワードマップ-岩手県
 (<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/03>) 2019年12月25日時点。

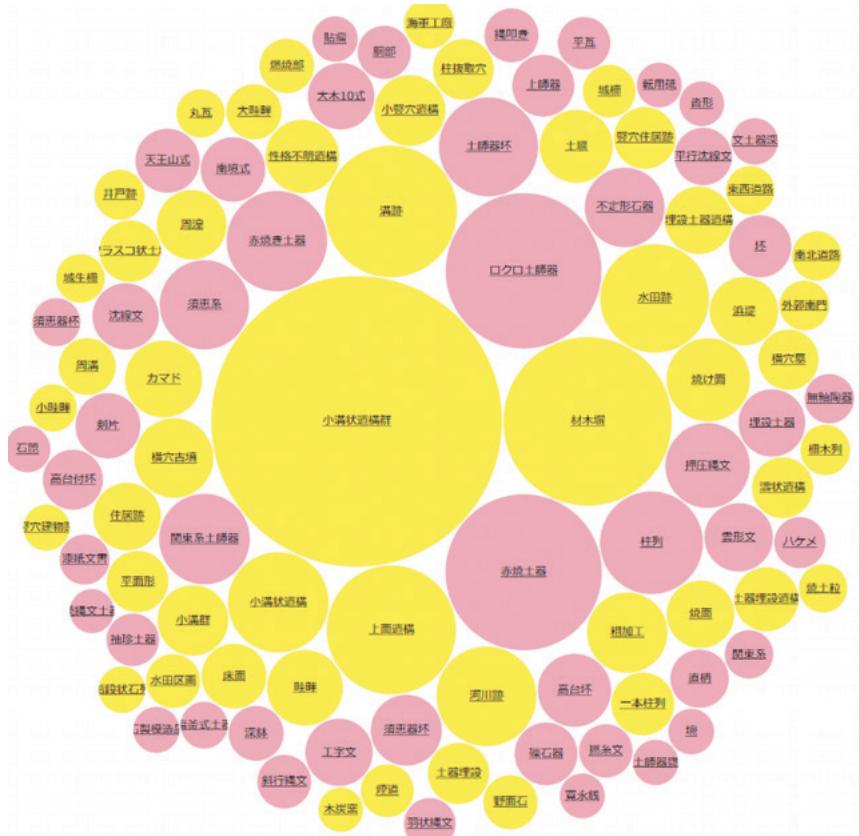


図4 報告書特徴語ワードマップ-宮城県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/04>) 2019年12月25日時点。

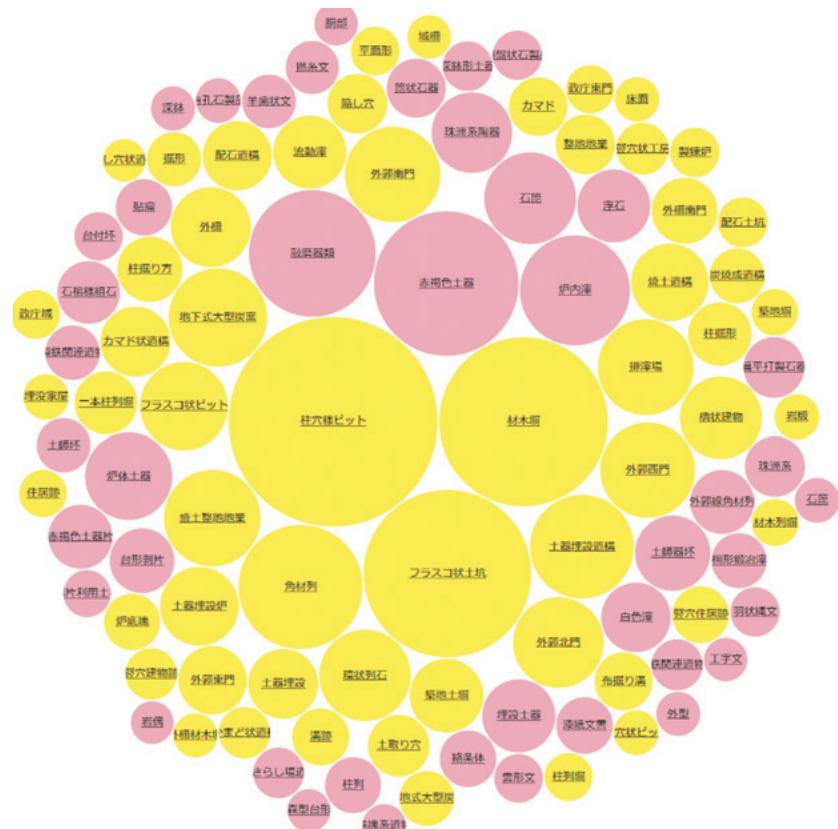


図5 報告書特徴語ワードマップ-秋田県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/05>) 2019年12月25日時点。

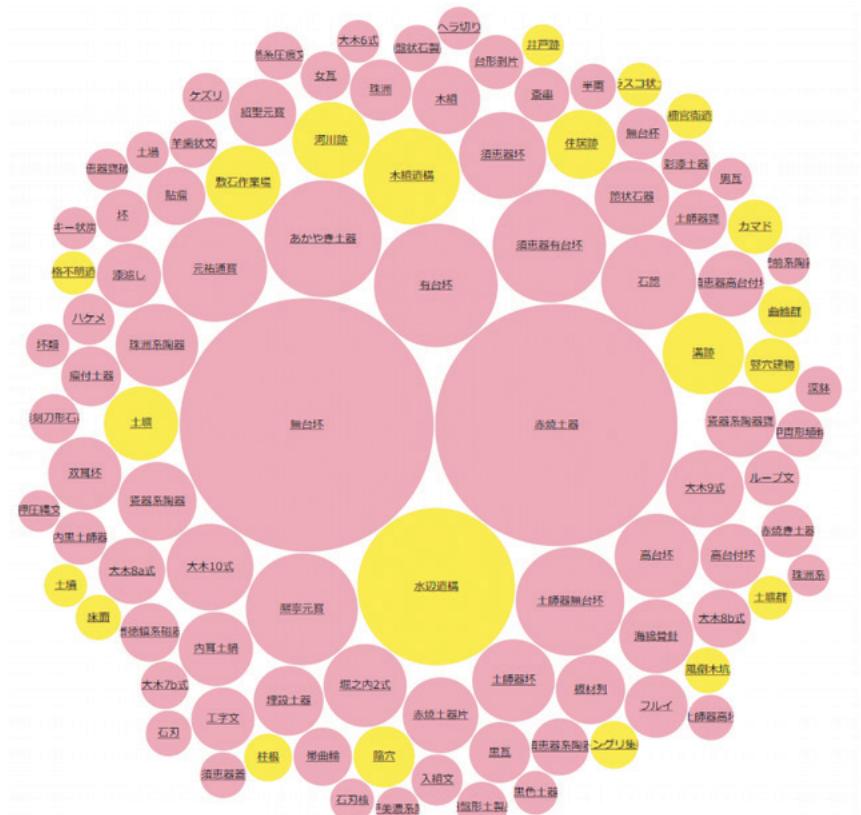


図6 報告書特徴語ワードマップ-山形県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/06>) 2019年12月25日時点。

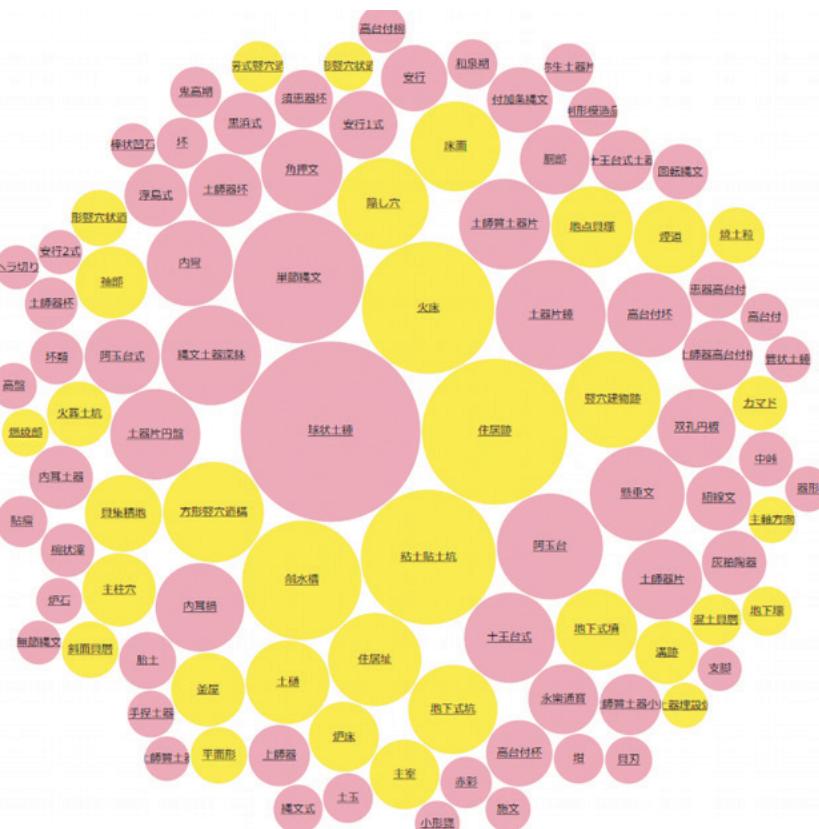


図7 報告書特徴語ワードマップ-茨城県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/08>) 2019年12月25日時点。

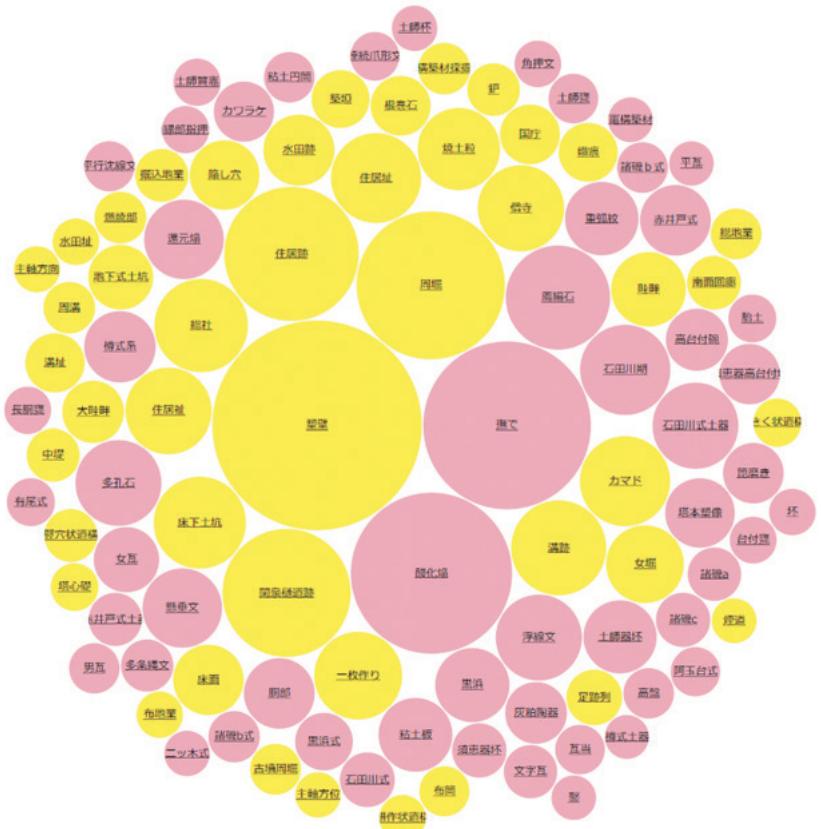


図8 報告書特徴語ワードマップ-群馬県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/10>) 2019年12月25日時点。

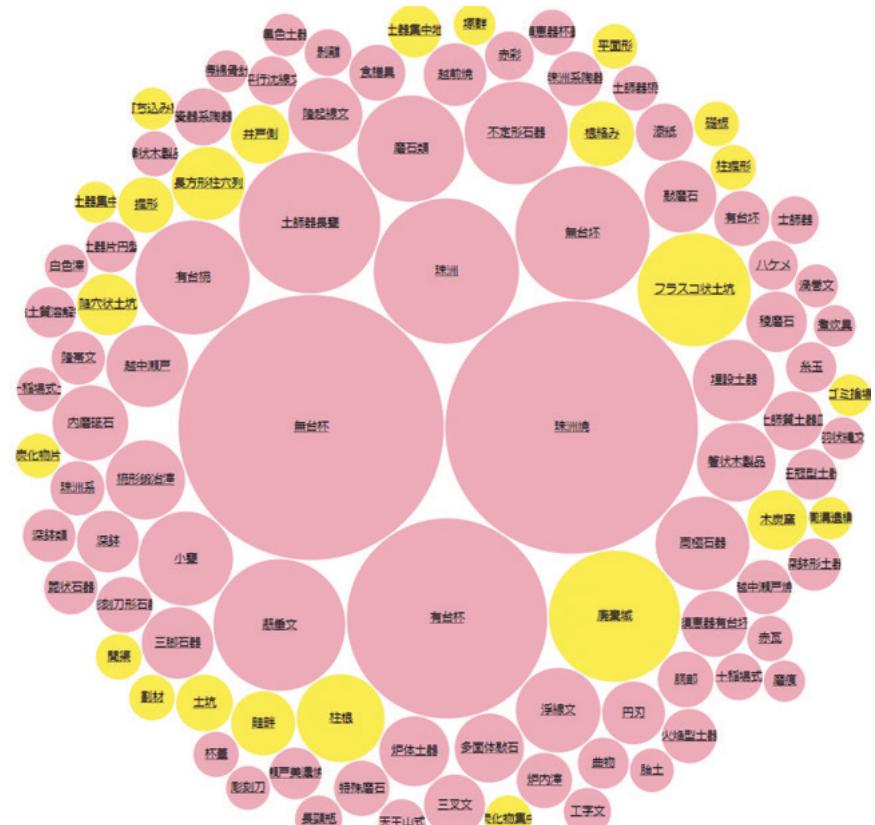


図9 報告書特徴語ワードマップ-新潟県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/15>) 2019年12月25日時点。

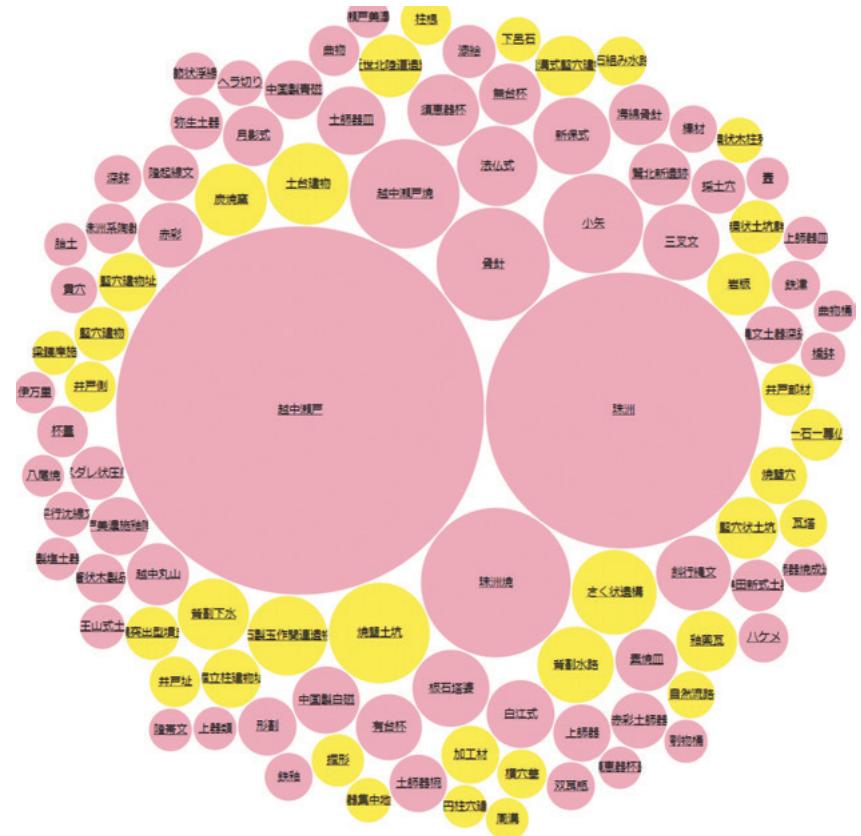


図10 報告書特徴語ワードマップ-富山県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/16>) 2019年12月25日時点。

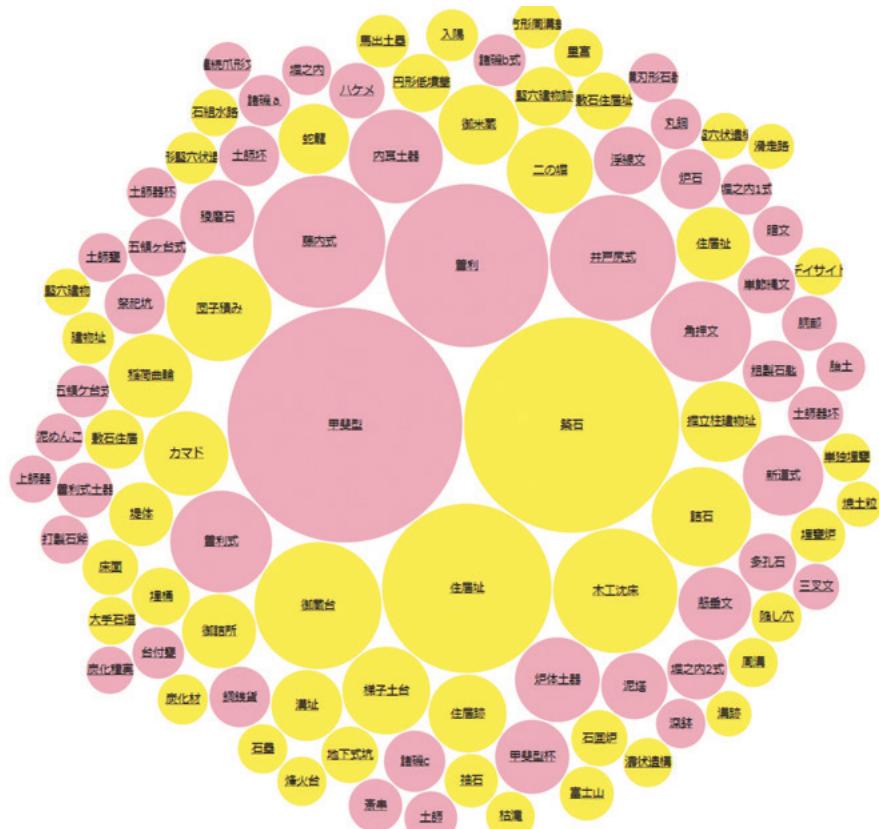


図11 報告書特徴語ワードマップ-山梨県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/19>) 2019年12月25日時点。

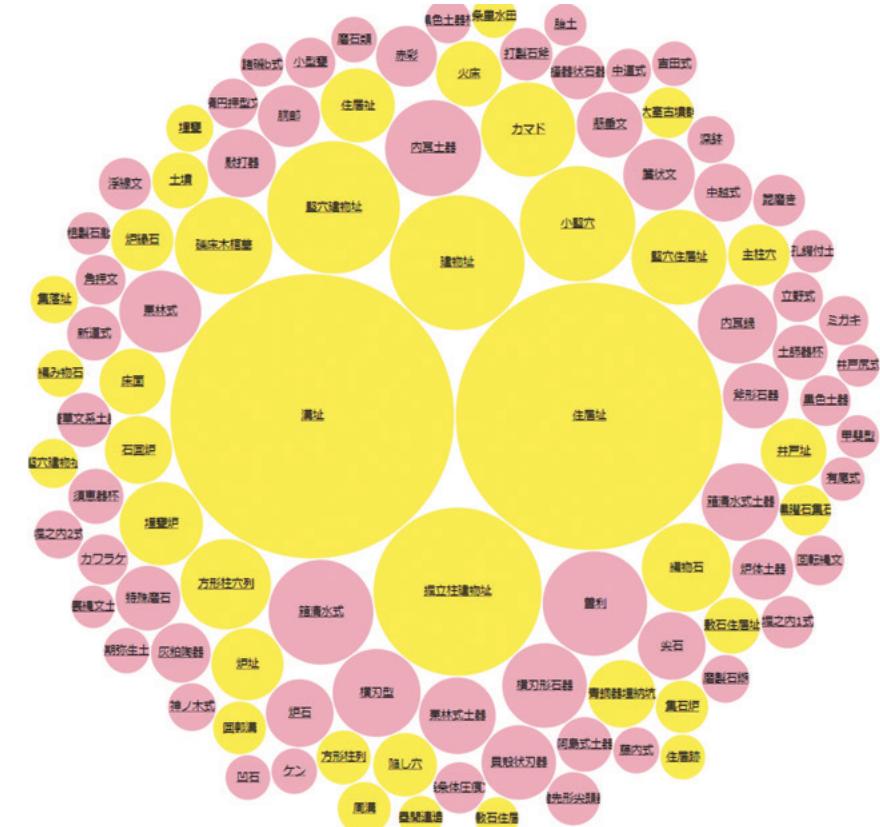


図12 報告書特徴語ワードマップ-長野県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/20>) 2019年12月25日時点。

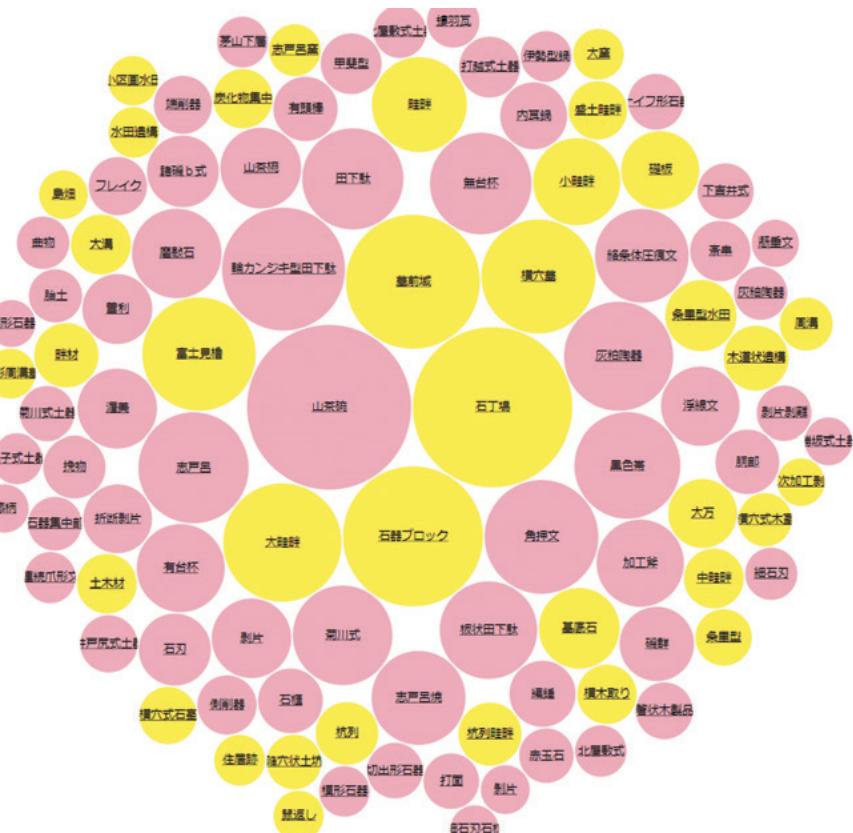


図13 報告書特徴語ワードマップ-静岡県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/22>) 2019年12月25日時点。

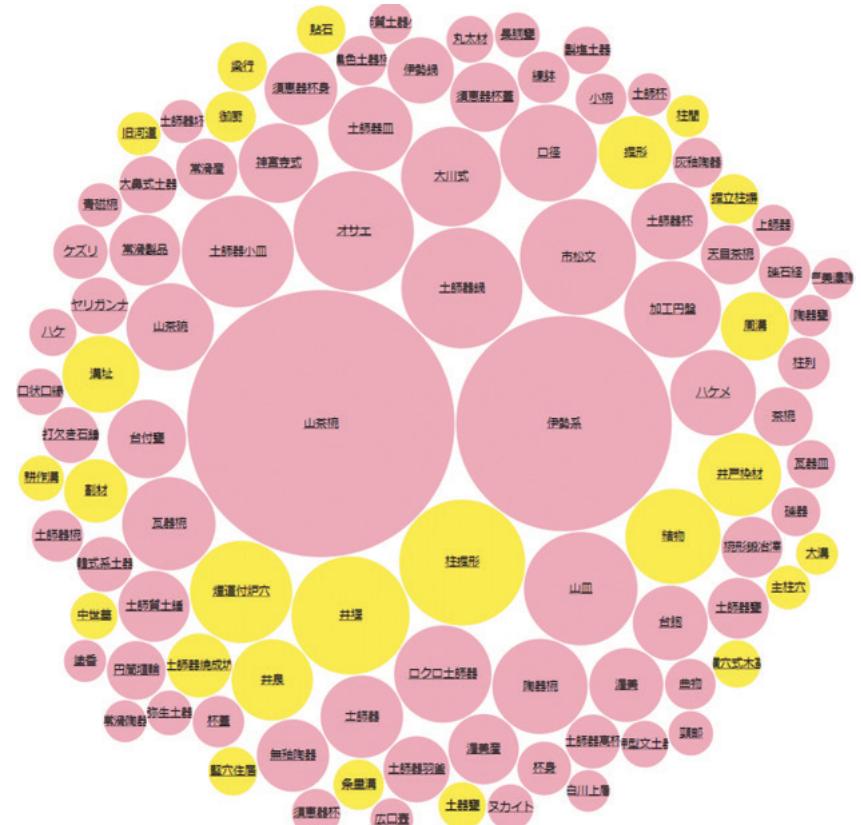


図14 報告書特徴語ワードマップ-三重県
 (<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/24>) 2019年12月25日時点。

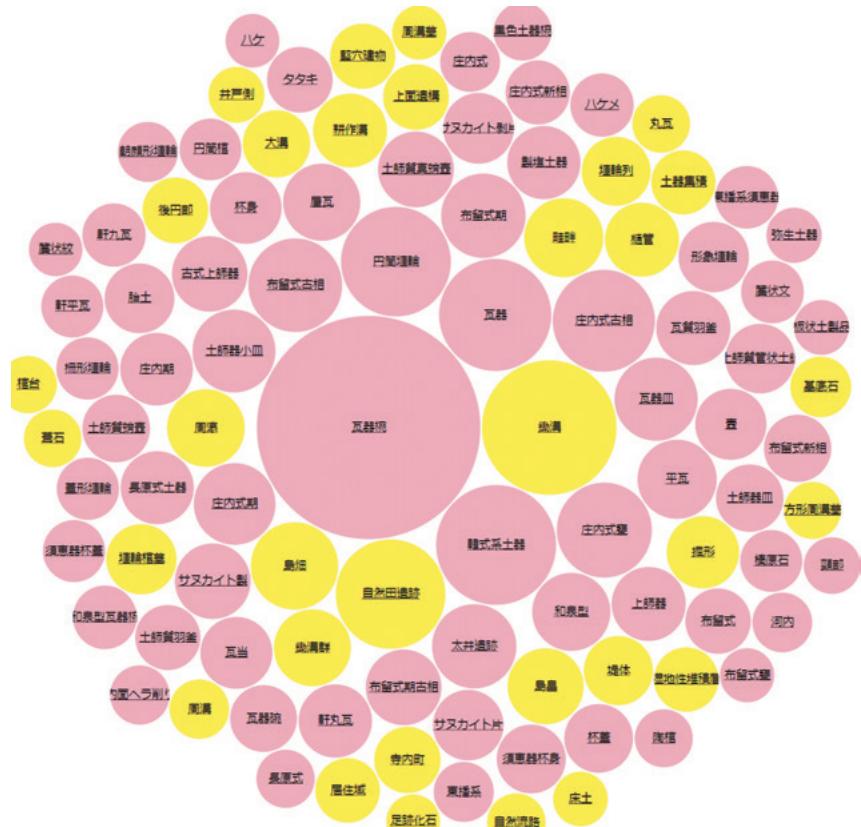


図15 報告書特徴語ワードマップ-大阪府
 (<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/27>) 2019年12月25日時点。

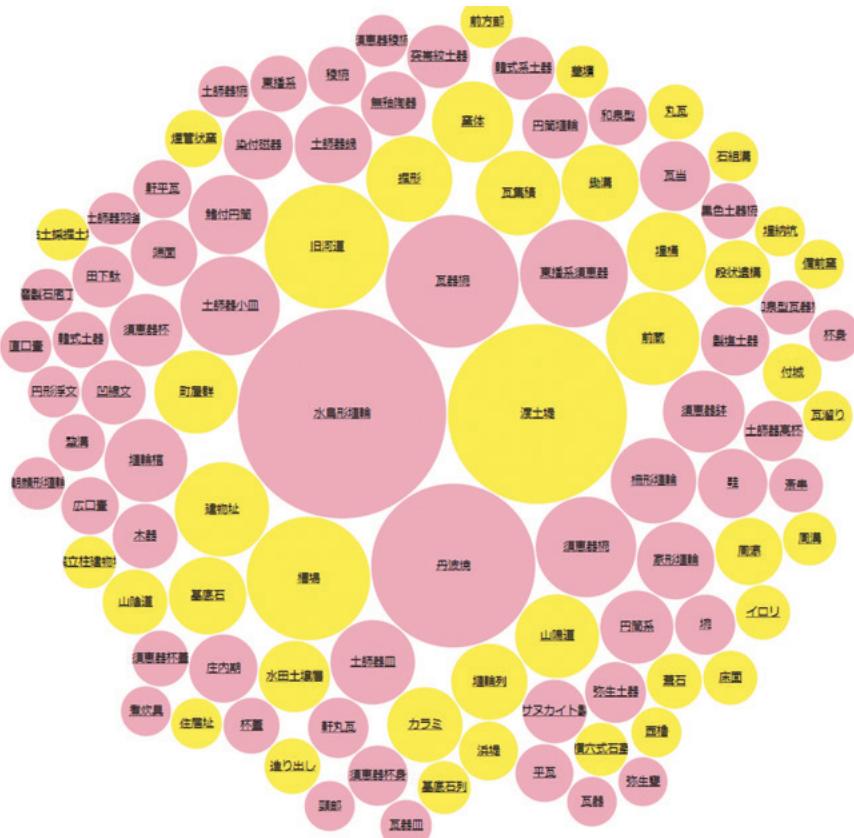


図16 報告書特徴語ワードマップ-兵庫県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/28>) 2019年12月25日時点。

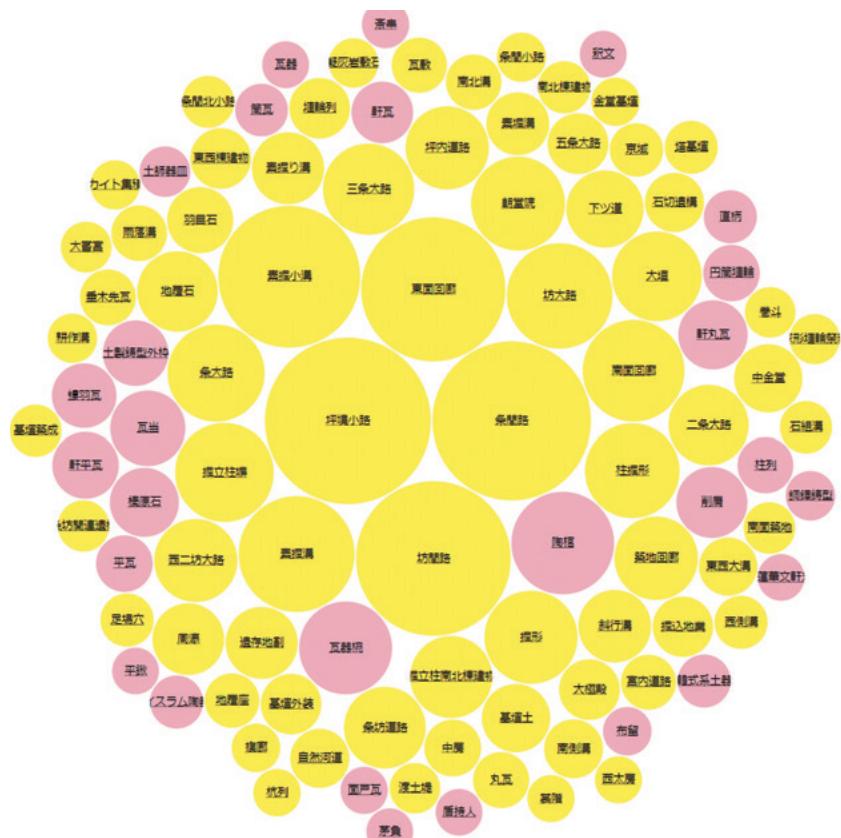


図17 報告書特徴語ワードマップ-奈良県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/29>) 2019年12月25日時点。

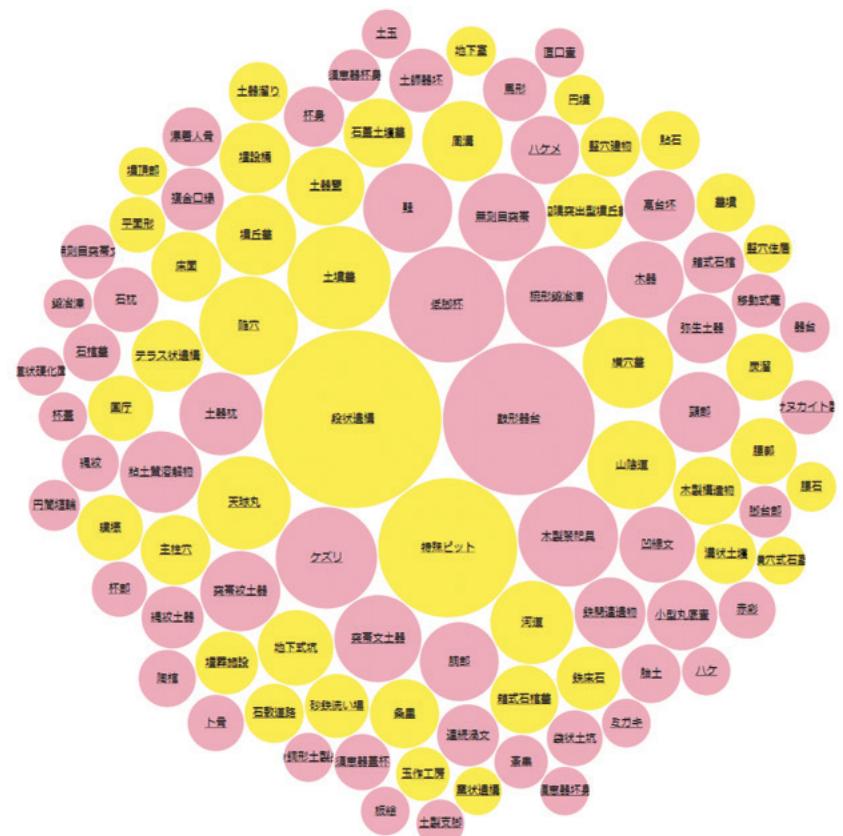


図18 報告書特徴語ワードマップ-鳥取県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/31>) 2019年12月25日時点。

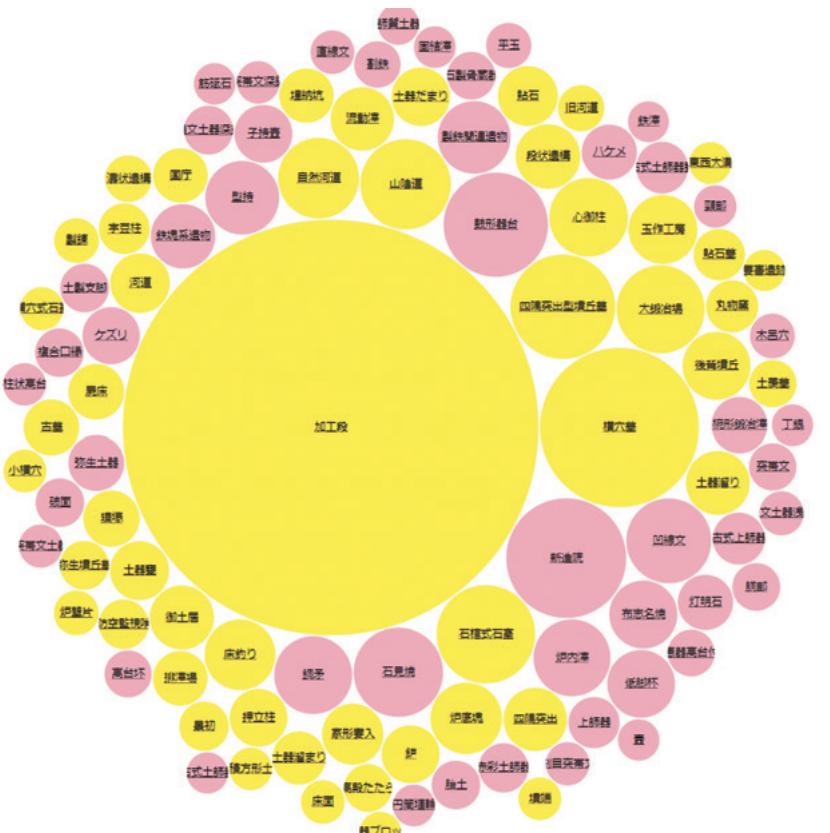


図19 報告書特徴語ワードマップ-島根県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/32>) 2019年12月25日時点。

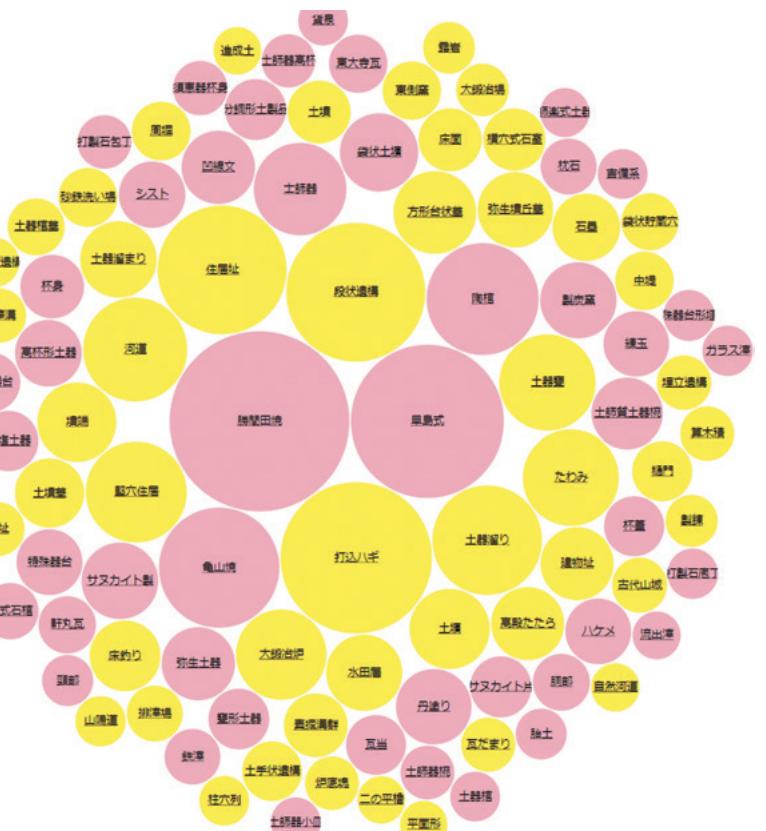


図20 報告書特徴語ワードマップ-岡山県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/33>) 2019年12月25日時点。

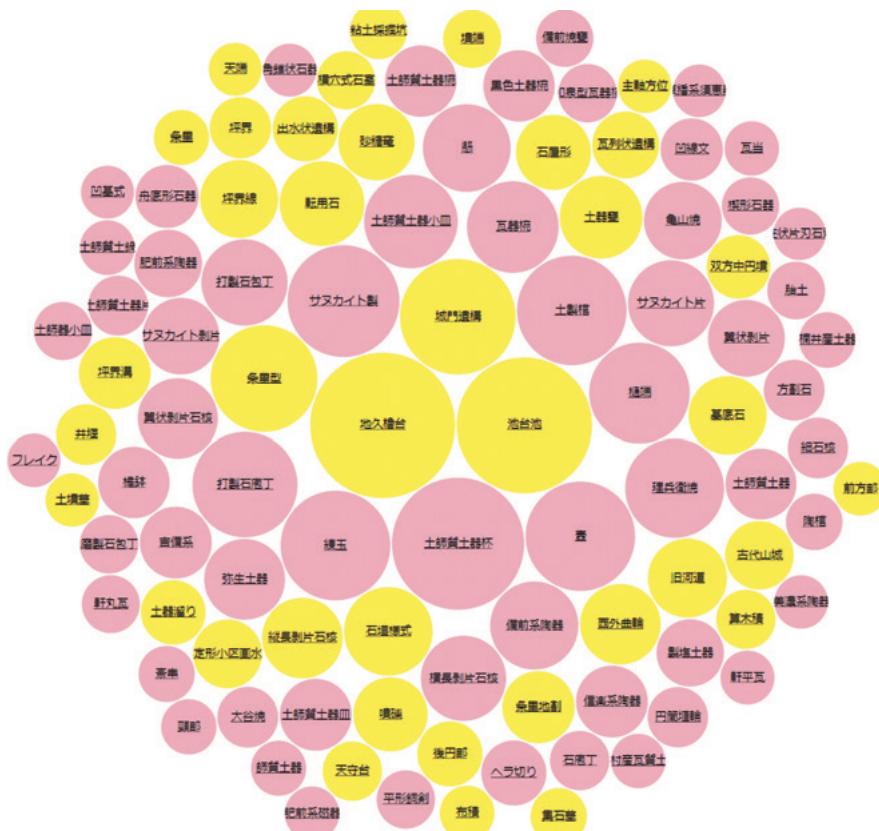


図21 報告書特徴語ワードマップ-香川県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/37>) 2019年12月25日時点。

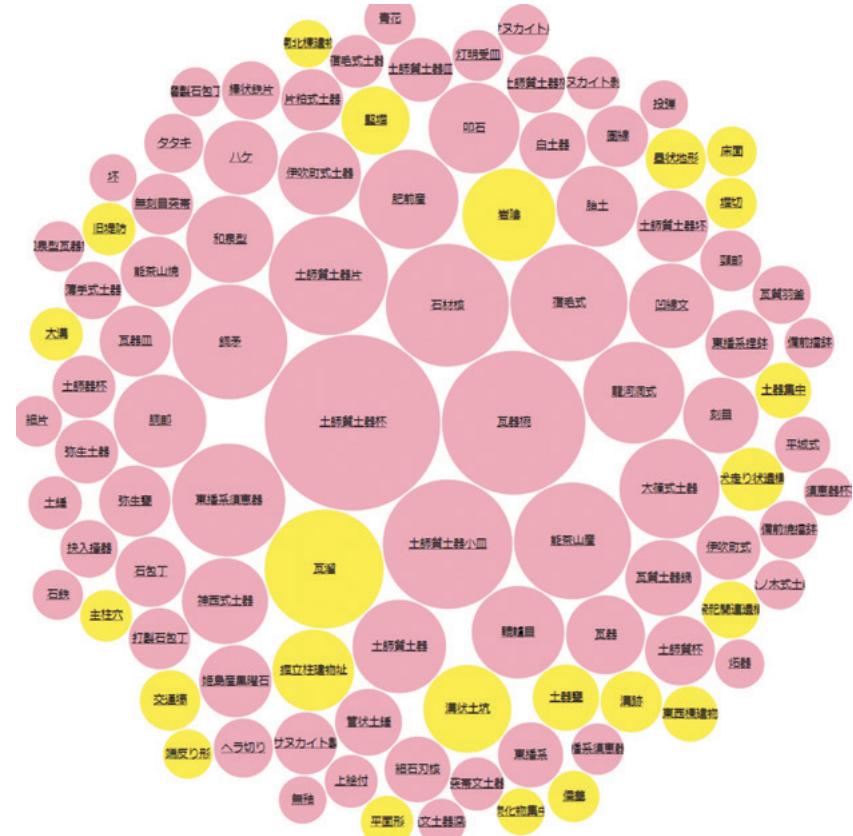


図22 報告書特徴語ワードマップ-高知県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/39>) 2019年12月25日時点。

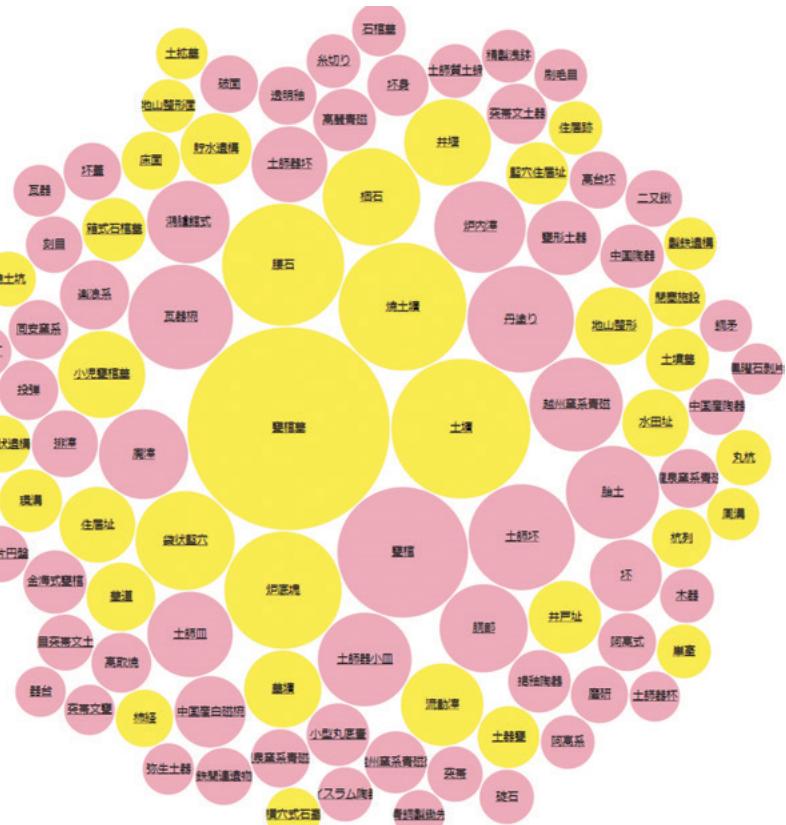


図23 報告書特徴語ワードマップ-福岡県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/40>) 2019年12月25日時点。

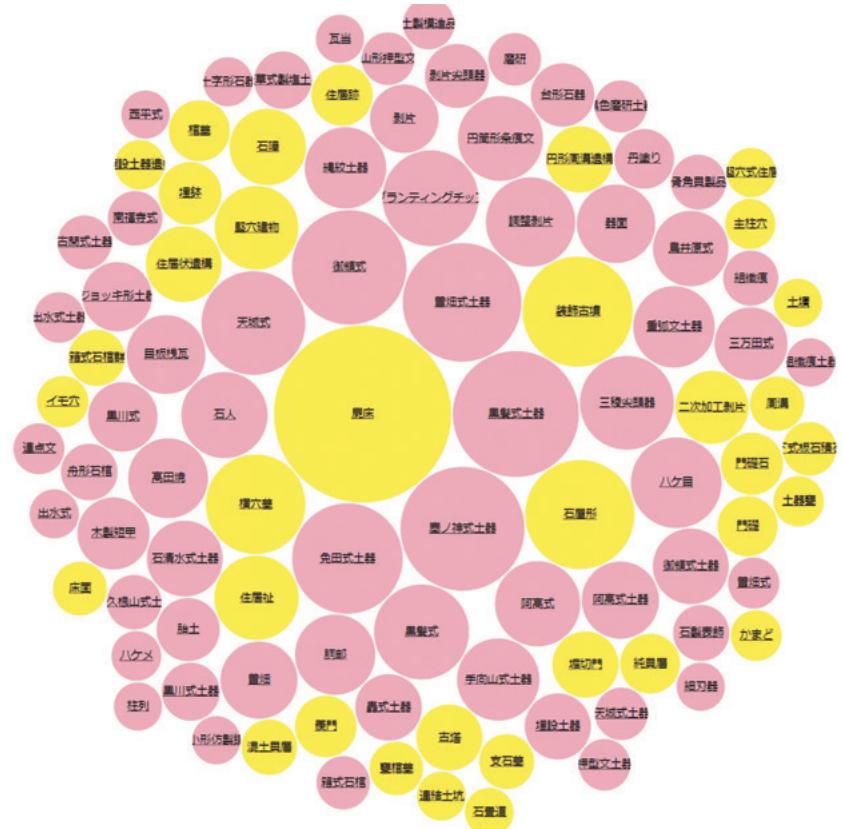


図24 報告書特徴語ワードマップ-熊本県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/43>) 2019年12月25日時点。

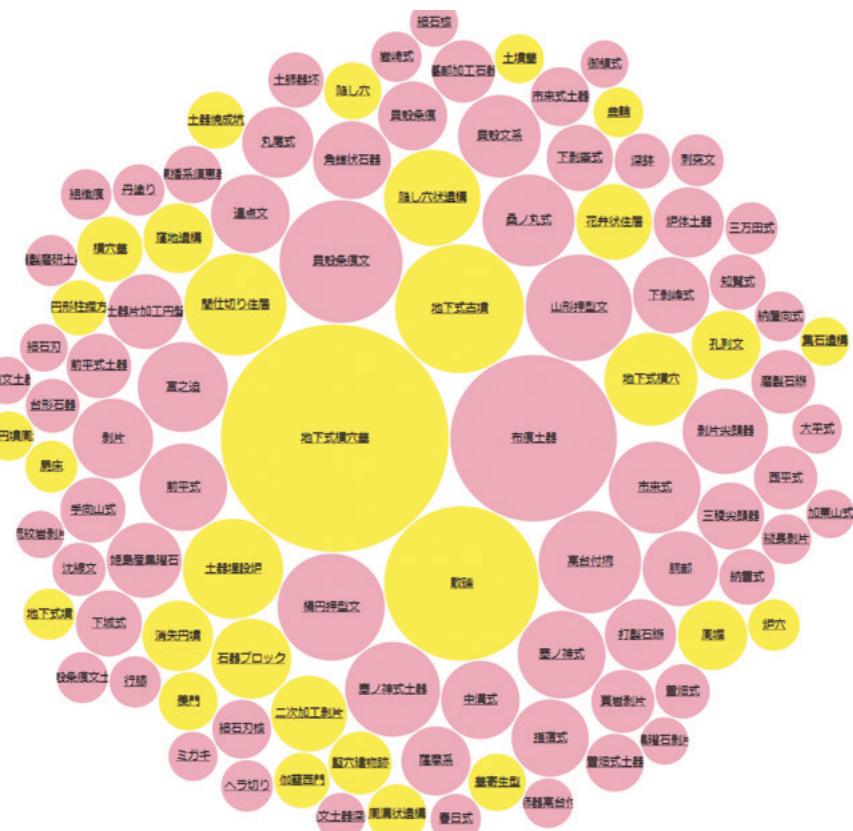


図25 報告書特徴語ワードマップ-宮崎県
(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/45>) 2019年12月25日時点。

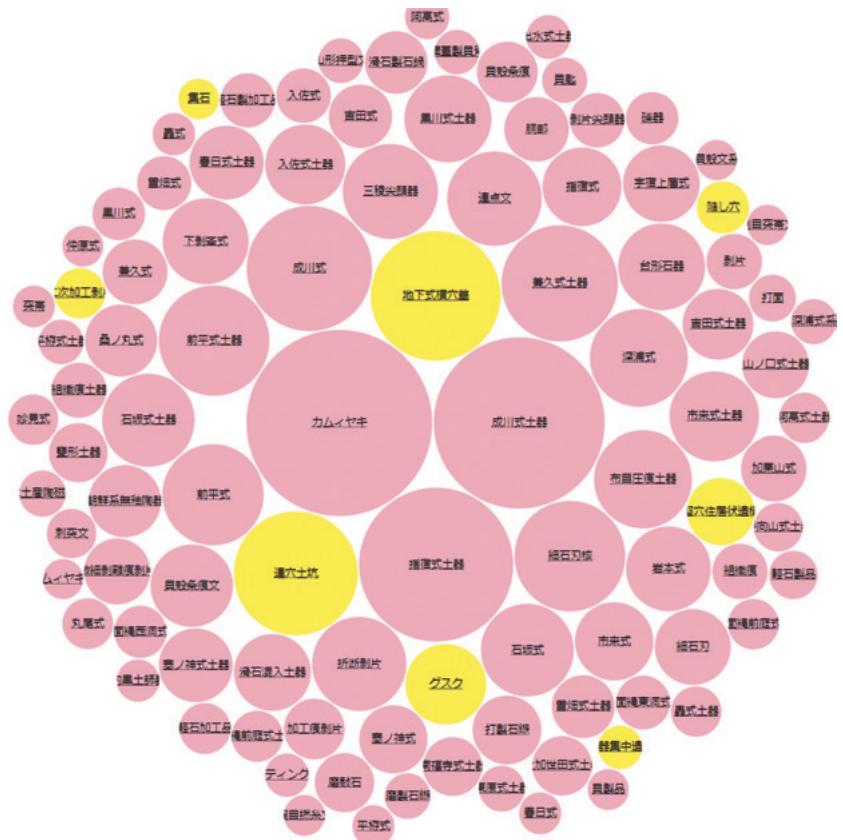


図26 報告書特徴語ワードマップ-鹿児島県

(<https://sitereports.nabunken.go.jp/ja/visualization/term/pref/46>) 2019年12月25日時点。