

# 一 般 競 争 入 札 公 告

科学技術・学術政策研究所において、下記のとおり一般競争入札に付します。

- 1 競争入札に付する事項
  - (1) 件 名 機関同定結果の検証用データセットの作成
  - (2) 請負期間 契約締結日から令和3年3月31日まで
  - (3) 納入場所 入札説明書のとおり
- 2 競争に参加する者に必要な資格
  - (1) 文部科学省競争参加資格（全省庁統一資格）において令和2年度に「役務の提供等」のA、B、C又はDの等級に格付けされ関東・甲信越地域の競争参加資格を有する者であること。
  - (2) 入札関係書類の提出時に、支出負担行為担当官が別に指定する暴力団等に該当しない旨の誓約書を提出した者であること。但し、支出負担行為担当官が誓約書の提出を要しないと認める場合は、この限りではない。
- 3 入札書等の提出場所等
  - (1) 入札関係書類の提出場所、契約条項を示す場所、入札説明書を交付する場所  
郵便番号 100-0013  
所在地 東京都千代田区霞が関3-2-2 中央合同庁舎第7号館東館16階  
機 関 名 科学技術・学術政策研究所総務課経理係  
電話番号 03-3581-2391
  - (2) 入札説明会の日時及び場所  
次の通り開催する。  
令和2年10月29日（木）11時00分  
科学技術・学術政策研究所大会議室（中央合同庁舎第7号館東館16B）
  - (3) 入札書及び入札関係書類の受領期限  
令和2年11月12日（木）12時00分
  - (4) 開札の日時及び場所  
令和2年11月19日（木）16時00分  
科学技術・学術政策研究所小会議室（中央合同庁舎第7号館東館16V）
- 4 入札保証金  
免除する。
- 5 入札の無効
  - (1) 本公告に示した競争参加に必要な資格のない者の提出した入札書、入札者に求められる義務を履行しない者の提出した入札書、その他文部科学省発注工事請負等契約規則第11条第1項各号に掲げる入札書は無効とする。
  - (2) 2（2）の誓約書を提出せず、又は虚偽の誓約をし、若しくは誓約書に反することとなったときは、当該者の入札を無効とする。
- 6 その他  
本件の入札に関する必要事項については、入札説明書によるものとする。

以上公告する。

令和2年10月20日

支出負担行為担当官  
科学技術・学術政策研究所長  
菱山 豊

# 仕 様 書

## 1. 業務題目

機関同定結果の検証用データセットの作成

## 2. 業務の目的

科学技術・学術政策研究所(以下、「NISTEP」という。)では、文部科学省「科学技術イノベーション政策における『政策のための科学』」推進事業におけるデータ・情報基盤整備の一環として、NISTEP 大学・公的機関名辞書(以下「機関名辞書」という。)の整備を進めるとともに、この機関名辞書を用いて、文献データベース(Web of Science Core Collection 及び Scopus)における国内機関名の同定(名寄せ)を実施している<sup>1</sup>。

この名寄せは、NISTEP で開発した機関同定プログラムを用いて行うが、データベース中の機関表記は同一の機関に対しても非常に多様であるため、プログラムによる同定が不正確あるいは不能である場合も起こり、データチェックが必要である。しかし、同定対象のデータ量が膨大であるため、目視によるデータチェックには専門家による多大の時間と労力を要している。

本業務の目的は、既存の同定データ(機関同定ができなかったデータを含む)からサンプルを抽出し、データの信頼性(同定が正しい確率)を推定する検証用データセットを作成することである。経験的には、高い信頼性のデータが大多数を占めると考えられるので、このデータセットができれば、目視チェック作業の負荷が大幅に低減すると期待される。

## 3. 業務の内容

受託者は、上記の目的を達成するため、以下の(1)～(5)を実施すること。

業務の実施にあたっては、NISTEP の担当者と定期的に打合せを行い、業務の進捗報告等を行うこと。

### (1) サンプルデータセットの抽出

受託者は、別途貸与する Web of Science Core Collection (以下、「WoSCC」という)(2020 年版)

---

<sup>1</sup> これまでの業務で得られた結果についての公表データ及び説明資料には以下のものがある。

1) NISTEP 大学・公的機関名辞書 ver2020.1 [http://doi.org/10.15108/data\\_rsorg001\\_2019\\_1](http://doi.org/10.15108/data_rsorg001_2019_1)

2) 大学・公的機関名表記ゆれテーブル ver2019.1 [http://doi.org/10.15108/data\\_rsorg002\\_2020\\_1](http://doi.org/10.15108/data_rsorg002_2020_1)

3) WoSCC\_NISTEP 大学・公的機関名辞書対応テーブル ver2017.1.2  
[http://doi.org/10.15108/data\\_rsorg003\\_2017\\_1](http://doi.org/10.15108/data_rsorg003_2017_1)

4) Scopus\_NISTEP 大学・公的機関名 P 辞書対応テーブル ver2018.1.1  
[http://doi.org/10.15108/data\\_rsorg004\\_2018\\_1](http://doi.org/10.15108/data_rsorg004_2018_1)

5) 小野寺夏生. 大学・公的機関における研究開発に関するデータの整備—マイクロデータ分析への貢献—. NISTEP NOTE No.11. 科学技術・学術政策研究所, 2014 年 5 月. <http://hdl.handle.net/11035/2926>

6) 小野寺夏生, 伊神正貴, 阪彩香. NISTEP 大学・公的機関名辞書の整備とその活用—大学下部組織レベルの研究データ分析に向けて—. NISTEP NOTE No.15. 科学技術・学術政策研究所, 2015 年 10 月.  
<http://hdl.handle.net/11035/3085>

7) 小野寺夏生, 伊神正貴, 富澤宏之. (2018). 客観的根拠(エビデンス)に基づく政策のためのデータ・情報基盤(第二回)～NISTEP 大学・公的機関名辞書～. STI Horizon, 4(3), 54-59,  
<http://doi.org/10.15108/stih.00147>

及び Scopus(2020 年版)の機関同定の結果から、それぞれ 25,000 件の同定データを抽出して、サンプルデータセットを作成すること。抽出は無作為ではなく、データベース上で出現数の多い機関・アドレス表記データの抽出度数を抑制し、出現数が 20 回以上の表記データが少なくとも 1 件抽出されるように調整すること。具体的な抽出方法は、NISTEP の担当者と協議すること。なお、抽出データには、データベースの記事番号、記事中のアドレス連番、論文発表年を含むこと。

(2) サンプルデータセットの機関表記データと機関名辞書の名称データとの類似度算出

受託者は、(1)より作成したサンプルデータセット中の各表記データと、別途貸与する機関名辞書(202005 改訂版)中の英語機関名データとの類似度を算出し、最も類似度の高い機関名辞書名称データとその機関名(日本語の正式名称)を記録すること。ある値以上の類似度を持つ機関が複数存在する場合は、それらのすべてを記録すること。類似度の算出は、NISTEP の同定処理プログラムで用いている N-Gram 文字列マッチとベクトルマッチの方式を基本とし、辞書内の複数の英語名表記のうち最大類似度でマッチした表記が分かるような方式とすること。類似度の算出方法についての詳細は NISTEP の担当者と協議して定めること。

(3) 類似度による同定データの目視チェック

受託者は、(2)により類似度が高いとされた機関名辞書中の機関(1 つの表記データに対し複数存在することがある)の正否を、別途貸与する WoSCC(2020 年版)及び Scopus(2020 年版)の機関同定の結果を参考にして判定すること。正解機関が不明の場合は、その旨記載すること。最高類似度の機関が正解機関と異なる場合には、可能な限り、両者の関係が以下のいずれであるか記録すること:(a)無関係の機関、(b)代表機関とその下位機関の関係、(c)同じ代表機関に属する下位機関の関係、(d)変遷前後の旧機関と新機関の関係。

(4) 検証用データセットの構築

以上の結果に基づき、受託者は、本委託事業の目的に沿う検証用データセットを作成すること。検証用データセットは、類似度(ある値以上の類似度を持つ機関が複数存在する場合は最高の類似度)の値によりデータを 3 ないし 4 段階に分け、各データについて、データベースの機関名表記データとその他の抽出データ(記事番号、記事中のアドレス連番、論文発表年)、マッチした機関名辞書の英語名称データと該当機関名、得られた類似度の値、正解と判定した機関名、最高類似度の機関が正解機関と異なる場合の両者の関係(可能な場合)を記載する。

4. 実施機関選定条件応募者に求める要求要件

- 特許、論文、ウェブ上のテキスト情報のいずれかを対象とした自然言語処理や機械学習等による分析について 5 年以上の経験を有する職員が在籍し、それらの職員のいずれかが本件を担当すること。
- 大規模データベースを用いた分析やそのデータマイニング及びそれらの結果の可視化といった極めて高い情報処理能力と実績を組織として持つこと。
- 科学技術分野の情報について、ウェブ上で情報検索が可能な知見を組織として持つこと。
- 本調査を円滑に実施するための人員体制を有すること。

## 5. 業務実施期間

契約日から令和3年3月31日

## 6. 納品物

委託業務の成果物として以下を電子媒体として提出すること(各1部)。なお、成果物について、本仕様書に合致していない部分については、必要に応じて追加または修正を求めるものとする。

(1) 3(4)により作成した検証用データセット

(2) 3の作業のため作成したプログラムとその解説書

上記(2)で作成するプログラムは Perl, Python, Ruby 等のスクリプト言語を使用すること。

## 7. 納入場所

東京都千代田区霞が関3丁目2番2号

文部科学省 科学技術・学術政策研究所 科学技術・学術基盤調査研究室

## 8. 無償貸付を行う資料

NISTEP は、受託者に対し、下記の資料の無償貸付を行う。

○ NISTEP 大学・公的機関名辞書(202005 改訂版)

○ WoSCC(2020 年版)の機関同定の結果

○ Scopus(2020 年版)の機関同定の結果

○ その他、必要に応じ本業務に関する平成 23 年度～令和元年度の成果物

## 9. 守秘義務

受注者は、本業務の実施で知り得た非公開の情報を如何なるものにも漏洩してはならない。

受注者は、本業務に係わる情報を他の情報と明確に区別して、善良な管理者の注意を持って管理し、本業務以外に使用してはならない。

## 10. その他

本仕様書に記載されていない事項、または本仕様書について疑義が生じた場合は、科学技術・学術政策研究所と適宜協議を行うものとする。

以上