# Is Tofu the Cheese of Asia?: Searching for Corresponding Objects across Geographical Areas

Yating Zhang*, Adam Jatowt and Katsumi Tanaka
Graduate School of Informatics, Kyoto University
Yoshida-honmachi, Sakyo-ku
606-8501 Kyoto, Japan
{zhang,adam,tanaka}@dl.kuis.kyoto-u.ac.jp

## ABSTRACT

Keyword-based search engines are widely used nowadays for content retrieval. Creating queries is relatively easy when users wish to retrieve content in familiar domains (e.g., information about things within their own country). However, they often struggle when searching in unfamiliar domains (e.g., searching for information related to a foreign country). In this paper, we approach the vocabulary gap problem by allowing users to search by analogical examples, that is, by letting them utilize information in familiar domains to perform search in domains unfamiliar to them. In particular, we focus on geographical domains. We propose to build connections between two different spaces (e.g., USA and Japan) by mapping the distributed word representations in one space with the ones in the other space. We first introduce an effective technique for automatically constructing seed pairs of terms to be used for finding the optimal mapping function. Then we propose general and topic-based transformations of terms from one space to another. We test the performance of the proposed approaches on datasets derived from Wikipedia which are related to two quite diverse countries: Japan and USA.

## CCS Concepts

•**Information systems** → **Question answering; Information extraction;** *Location based services;*

## Keywords

spatial counterpart, object search, spatial transformation

*New affiliations: RIKEN Center for Advanced Intelligence Project (AIP), Japan
Nara Institute of Science and Technology, Japan
E-mail address: yatingz89@gmail.com


*WWW 2017 Companion*, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3041021.3055132

## 1. INTRODUCTION

Keyword-based retrieval is very common these days. It is based on an implicit assumption that searchers are relatively familiar with domains of their search. However, users often need to search in unfamiliar domains such as when seeking for information related to different spatial areas. Take as an example, Michael from USA who is a fan of *New York Yankees* and is visiting Japan. Suppose that he wants to know the name of a similar team in Japan. The Japanese counterpart of "New York Yankees" such as `Yomiuri Giants` could be then a possible answer. In another example the user may want to search for information about "Japanese NASA". `JAXA` is the term that should be suggested in this case. Analogously, `typhoon` in Japan would be the closest concept to "hurricane" in USA.

To fill in the knowledge gap between the areas that the user has much knowledge about and target areas which are unfamiliar to her (e.g., foreign countries), we propose automatic translation mechanisms. They will allow users to search by analogical objects as in the examples mentioned above. In this paper, in particular, we focus on the problem of translation across geographic places such as different countries. Searching about concepts specific to foreign countries or areas occurs relatively commonly these days. In the current era of globalization and frequent, free travel, users often need to perform this kind of search, for example, during traveling abroad or when interacting with foreigners.

Formulating queries in the form of analogy helps searchers avoid giving detailed descriptions of searched objects (e.g., Michael would have to come up with good descriptive phrases to describe his search intent). In fact, it may be difficult or impossible for users to describe desired objects in a concise form. Furthermore, user-provided descriptions do not offer guarantee that the desired entities will be actually returned. On the other hand, querying by analogical examples is relatively easy. This kind of search can be seen as a process that automatically enriches searcher's knowledge by constructing connections between objects across different spaces and it can be also regarded as a type of associative reasoning.

Detecting analogical objects in heterogeneous geographical spaces is however not trivial. The biggest challenge lies in the difference between contexts of different spatial areas[1], suggesting that the direct context comparison will not work. To solve this issue, for each spatial area, we first utilize distributed word embedding technique [14, 15] to decrease

[1]The differences result from different cultures, environment, history, etc.

the dimension of word meaning representation (conceptually portrayed in Fig.4). Then, given the two distributed vector spaces, we align them by training mapping functions. In result, based on the established alignment we can later retrieve the list of top counterparts.

However, the inherent problem behind such an approach is the difficulty of finding large training sets for training the mapping function across two spaces given the variety of aspects, topics, document genres in typical realistic scenarios. Thus to design a robust method we propose using automatically derived training sets to construct transformation matrices for a given pair of two spatial areas (e.g., Japan as a base area and USA as the target area). We utilize the *shared concepts* as seeds to train the mapping function (or as anchors to align two vector spaces) by assuming that general terms are more likely to be shared by different spatial areas than more specific terms. Once the spatial transformation framework is set up, we introduce several methods for detecting spatial counterparts. They perform several topic-biased transformations, the results of which are then merged for selecting the most corresponding object in another geographical space.

To sum up, our contributions are as follows:

1. We propose efficient and effective framework to search for analogical objects across spatial areas based on general and topic-biased transformations.
2. We evaluate the proposed approaches on the Wikipedia's unstructured text, which prove the effectiveness of our approach.
3. An important characteristic of our approach is that it is unsupervised. The proposed methods are also generic enough to be applied for any raw-text datasets.

## 2. PROBLEM STATEMENT

In this section, we formally define the problem of the across-space analogical object detection.

We set two spaces: a base space $S^b = \{w_1^b, w_2^b, ..., w_m^b\}$ ($w_i^b \in$ Vocabulary of $S^b$) from which the query is selected, and a target space $S^t = \{w_1^t, w_2^t, ..., w_n^t\}$ ($w_i^t \in$ Vocabulary of $S^t$) where the answer is to be retrieved from.

*Across-Space Analogical Object* is defined as an object $w^t$ (e.g., `Yomiuri Giants`) which is contextually similar to the queried object $w^b$ (e.g., *New York Yankees*). Note that the context between $w^t$ and $w^b$ is not required to be literally same. The literal form could be different as long as their meanings are similar.

## 3. GENERAL TERM TRANSFORMATION

Our goal is to compare terms related to disjoint geographical areas and to find matching term pairs (e.g., `NASA` and `JAXA`). For this, we propose constructing a mapping function between the base space and the target space. This process is query independent and can be done offline. While establishing the mapping function would necessarily require a supervised method, we assume in this work an unsupervised approach. This is because it is infeasible to provide sufficient number of training pairs of terms (such as the examples listed above) for any possible combination of two different countries or other geographic regions. We then resort to automatically finding training pairs. One way to generate such term pairs could be based on the equality of term literal forms. However, we cannot always assume a direct semantic correspondence between the same term in two different spaces. Even if the same term appears in two different spaces, there is no assurance that it indeed denotes identical concept. For example, `sushi` in Japan is regarded as typical or local food, however in another country, such as USA, though `sushi` also exists, the position/role behind it is rather different (e.g., `sushi` is regarded as foreign and relatively luxury food outside of Japan). This phenomena can be interpreted as the meaning shift across spaces. On the other hand, sometimes literally different terms in different spatial areas may represent the same or very similar concept, such as `haiku` in Japan and `poetry` in USA.

### 3.1 Word Embedding

For capturing and representing term semantics we apply word embedding techniques. Distributed representation of words by using neural networks was originally proposed in [17] and was later improved by Mikolov *et al.* [14, 15] who introduced Skip-gram model relying on a simplified neural network architecture for generating vector representations of words from text.

### 3.2 Transformation based on Anchor Mapping

The objective of our approach is to measure similarity between terms in the base space and terms in the target space for finding spatial analogs (which are called here also spatial counterparts). As we mentioned before, we cannot directly compare terms in two different semantic vector spaces. The reason is that the features/dimensions in both the spaces have no direct correspondence as a result of separate training processes. We then train a transformation matrix to establish the connection between the vector spaces.

To better understand the concept behind the transformation, one could compare the semantic spaces to buildings. If two semantic spaces are imagined as two buildings, to map the components of one building to the ones in the other one, we need first to learn how the main frames of the two buildings correspond to each other. After this is known, the remaining components can be automatically mapped by considering their relative positions to the main frames of their buildings. So, in our case, once the correspondence between the anchor terms in the two semantic spaces is established, one can automatically map any terms relative to these anchor terms. In this work, we propose to use *Shared Frequent Concepts* (SFC) as anchors to construct the transformation. However manually preparing large enough sets of anchor terms that would cover various topics/domains as well as exist in any possible combinations of the base and target spaces requires much effort and resources. We rely here on an approximation procedure for automatically finding SFC to be used as anchor pairs. Specifically, we select terms that (a) are general in their meaning and (b) have high frequency (e.g., mountain, river, lake, president) in both the base and the target spaces. The intuition behind this idea is that general terms that are also frequent in both spaces are more likely to have stable meaning and be also co-occurring with many other terms. The details of extracting SFC are described in the experimental settings (see Sec. 5.1).

Suppose there are $u$ pairs of anchor terms $\{(x_1^b, x_1^t),\ldots,(x_u^b, x_u^t)\}$ where $x_i^b$ is an anchor in one space (e.g., Japan) and $x_i^t$ is its counterpart, that is, the same anchor in the other space (e.g., USA). The transformation matrix $\mathbf{M}$ is then constructed by minimizing the differences between $\mathbf{M}\mathbf{x_i^b}$ and $\mathbf{x_i^t}$

(see Eq. 1). In particular, we minimize the sum of Euclidean 2-norms between anchor terms in the base space and their counterpart anchors in the target space. Eq. 1 is applied for solving the regularized least squares problem ($\gamma = .02$) together with regularization component for preventing overfitting:

$$\mathbf{M} = \underset{\mathbf{M}}{\operatorname{argmin}} \sum_{i=1}^{u} \left\| \mathbf{M}\mathbf{x_i^b} - \mathbf{x_i^t} \right\|_2^2 + \gamma \left\| \mathbf{M} \right\|_2^2 \tag{1}$$

$u$ denotes here the size of anchor term set which contains, in our implementation, the top 5% frequent concepts (over 10,000 terms) in the intersection of vocabularies of the two corpora.

Note that our approach is generic. For other types of heterogeneous spaces we need to train a dedicated transformation matrix to align the two vector spaces. The anchor terms in such a case should be terms having the same or very similar meaning in both the different spaces (e.g., frequent terms judged to represent the same concepts). On the other hand, if the similarity estimation task is conducted over the same vector space, then there is no need for transformation.

## 3.3 Retrieval Model for General Transformation

After obtaining the transformation matrix $\mathbf{M}$, we can compute the similarity of a query, $q$, in the base space with any term $v$ in the target space by multiplying the query's vector representation with the transformation matrix $\mathbf{M}$, and then by calculating the cosine similarity between such transformed vector and the vector $v$.

$$S_{sim}(\mathbf{q}, \mathbf{v}) = cos(\mathbf{Mq}, \mathbf{v}) \tag{2}$$

Terms that have the highest similarity value are returned as answers. The similarity computation is illustrated in Fig. 1.

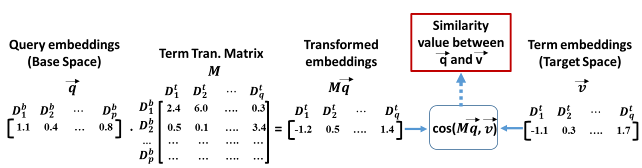


**Figure 1: Computing similarity between query $q$ in the base space and a term $v$ in target space using general term transformation.**

# 4. TOPIC-BIASED TRANSFORMATION

In Sec. 3, we explained how to align two spaces by training a transformation matrix. However, such an approach has implicit assumption that *the whole vocabulary in a given space is subject to the same transformation*, which is obviously crude. In this section, we introduce a way to train the transformation matrix biased on specific topics, called Topic-biased Term Transformation (**TT**). The motivation behind this approach is that *terms belonging to different topics should be subject to different transformations.* In Sec. 4.1 we first introduce the topic model used to extract the thematic structure of documents and to estimate the term distribution of a given topic as well as the topic distribution of a given term. In Sec. 4.2 and 4.3, we propose two variants of methods based on different hypotheses.

## 4.1 Topic Discovery

Latent Dirichlet Allocation (LDA) [4], one of the probabilistic topic models, has been successfully used for extracting hidden themes in large collection of documents. The idea behind LDA is to model each document as a mixture of topics, where a *topic* is defined to be a distribution over a fixed number of terms (vocabulary of the corpus) [4, 3]. LDA is a generative model in which the topic distribution is assumed to have a Dirichlet prior. Given the only observable variables - terms in documents belonging to the corpus, a topic model is trained to infer the hidden thematic structure behind the observations.

After learning is completed, the probability of a term $w$ to belong to a topic $z_k$ ($k \in [1, K]$), $P(w|z_k)$, is known[2]. Then the probability of a topic $z_k$ given a term $w$ can be easily inferred by applying Bayes' rule, $P(z_k|w) \propto P(w|z_k)P(z_k)$, where $P(z_k)$ is approximated by the exponential of the expected value of its logarithm under the variational distribution [4]. Therefore, through the LDA model, we can obtain the probabilistic distribution of topics given a term in the corpus. We use this information to determine (1) the importance (weight) of the anchors when training the transformation matrix for a specific topic (called topic-biased transformation matrix); (2) the topic-preference of a query to choose which (or to what extent) topic-based transformation matrices to be used for doing mapping. The details are described in the following sections.

## 4.2 Transformation under Same Topic Distribution across Spaces

As discussed in the beginning of Sec. 4, we aim at training transformation matrices for different topics. Our idea is to adjust Eq. 1 by involving a parameter $\phi$, expressed as the probability of an anchor $x_i$ given a topic $z_k$, $P(x_i|z_k)$, to bias the optimization of transformation matrix using anchors related to a given topic. Since we have two spaces, in other words, two collections of documents, then in order to calculate $P(x_i|z_k)$, we face the problem of whether the anchors in the two corpora share the same topic distribution or not. In this section, we test the first possibility by setting up the following hypothesis.

**Hypothesis 1**: *The anchors used for alignment share the same topic distribution across two spaces.*

To implement this idea, we combine the documents of the base space and the target space. Then we train one topic model based on such joint corpus. We next compute the probability of anchor terms given each topic, $P(x_i|z_k^{b,t})$, using the approach described in Sec. 4.1[3]. The probabilistic distribution of anchors in each topic ($\boldsymbol{\phi_k^{b,t}}$) is involved in training topic-based transformation matrix $\mathbf{M_k}$ on topic $k$ (k∈[1,K]). The optimization function for training the topic-based spatial transformation is expressed in Eq. 3. The higher the value of $\phi_{i,k}^{b,t}$ is, the more biased becomes the transformation matrix to topic $k$ when trained using the anchor term $x_i$.

$$\mathbf{M_k} = \underset{\mathbf{M_k}}{\operatorname{argmin}} \sum_{i=1}^{u} \phi_{i,k}^{b,t} \left\| \mathbf{M_k}\mathbf{x_i^b} - \mathbf{x_i^t} \right\|_2^2 + \gamma \left\| \mathbf{M_k} \right\|_2^2 \quad \text{where } \phi_{i,k}^{b,t} = P(x_i|z_k^{b,t}) \tag{3}$$

[2] $K$ is set to 20 in our experiments.

[3] We denote the values based on the joint corpus by superscript "b,t".

Note that in Eq. 3, the anchors' probability distribution given a topic, $P(x_i|z_k^{b,t})$, is normalized to sum to 1.

### 4.2.1 Retrieval Model for Topic-biased Transformation

We demonstrate the way to compute the similarity of a query $q$ in the base space with a term $v$ in the target space in Fig. 2. Since the two spaces share the same topic distribution, the similarity between query $q$ vector (shown in the left side of Fig. 2) and the vector of a term $v$ (shown on the right) can be obtained by taking the weighted sum of the cosine similarities between the topic-bases transformations of the query vector (the transformed vectors are shown in the center of the figure) and the vector of $v$ in the target space. The topical probabilities of the query (marked by red color in Fig. 2) are used here as weights.

$$S_{sim}(\mathbf{q},\mathbf{v}) = \sum_{k=1}^{K} \lambda_{q,k} \cdot cos(\mathbf{M_k q}, \mathbf{v}) \quad \text{where } \lambda_{q,k} = P(z_k^{b,t}|q) \tag{4}$$

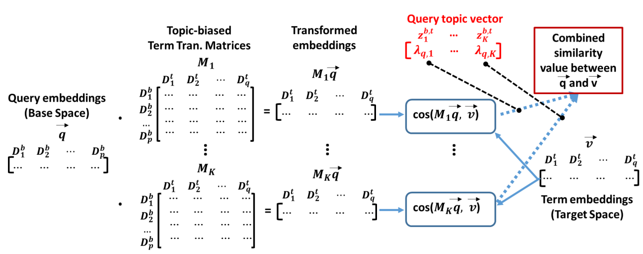


**Figure 2: Computing similarity between query $q$ in the base space and term $v$ in target space using topic-biased term transformation.**

## 4.3 Transformation under Different Topic Distribution across Spaces

In this section, we propose an alternative method to the one shown above. This method is particularly useful when the topic distributions are different in each corpus. For example, two countries are in quite different geographical environments (e.g., one is a coastal country and the other is a landlocked country), or the two spatial areas have quite different cultures, social activities, etc. Under this scenario, the Hypothesis 2 is applied:

**Hypothesis 2**: *The anchors used for alignment have different topic distributions across two spaces.*

As now the topic distribution given an anchor varies across spaces, there is no correspondence between the parameter $P(x_i|z_k^b)$ of anchor $x_i^b$ and $P(x_i|z_k^t)$ of its counterpart $x_i^t$. Simply speaking, for the two topic models trained on the base space and the target space separately, the theme of $k$th topic in the base space is different from that of $k$th topic in the target space. In such sense, before we transform term representations as shown in the previous sections (Eq. 1, Eq. 3), we have to first transform topics so as to also find the correspondence between the topic distributions across spaces. The overview of this process is also shown in Fig. 3.

### 4.3.1 Establishing Topic Transformation

To find the correspondence between topics across two corpora, we follow similar way as when transforming word embeddings. We train a topic transformation matrix $\mathbf{H}$ based on topic representations.

We use $\mathbf{x_i^{Z^b}}$ as a topic representation vector of anchor term $x_i$ in the base space, $[P(z_1^b|x_i), ..., P(z_K^b|x_i)]$. The probabilities of topics given a term are normalized and sum to 1. Similarly, in the target space, $\mathbf{x_i^{Z^t}} = [P(z_1^t|x_i), ..., P(z_K^t|x_i)]$. Note that the number of topics in the base space can be different from the number of topics in the target space, however, for the ease of explanation we use the same topic number $K$. Then the topic transformation matrix (not to mistake with topic-biased transformation matrix introduced before) $\mathbf{H}$ of dimension $K \times K$ can be trained in a similar manner of training the term representation transformation matrix using the same set of seeds. The optimization function is given in Eq. 5.

$$\begin{aligned} \mathbf{H} &= \underset{\mathbf{H}}{\operatorname{argmin}} \sum_{i=1}^{u} \left\| \mathbf{H x_i^{Z^b}} - \mathbf{x_i^{Z^t}} \right\|_2^2 + \gamma \left\| \mathbf{H} \right\|_2^2 \\ \text{where } \mathbf{x_i^{Z^b}} &= [P(z_1^b|x_i), ...P(z_k^b|x_i), ..., P(z_K^b|x_i)] \\ \mathbf{x_i^{Z^t}} &= [P(z_1^t|x_i), ...P(z_k^t|x_i), ..., P(z_K^t|x_i)] \end{aligned} \tag{5}$$

The topic transformation matrix is used to convert the topic distribution of query $q$ in the base space to its corresponding topic distribution in the target space (see Eq. 7).

### 4.3.2 Establishing Term Transformation Matrix

Since the two spaces now have different topic distributions and also because the user is searching in the target space, we utilize the probabilistic distribution of anchors in the target space ($\boldsymbol{\phi_k^t}$) to train the topic-biased transformation matrices. In other words, each of the trained matrix, $\mathbf{M_k}$, is biased on the specific topic in the target space (see Eq. 6).

$$\mathbf{M_k} = \underset{\mathbf{M_k}}{\operatorname{argmin}} \sum_{i=1}^{u} \phi_{i,k}^t \left\| \mathbf{M_k x_i^b} - \mathbf{x_i^t} \right\|_2^2 + \gamma \left\| \mathbf{M_k} \right\|_2^2 \quad \text{where } \phi_{i,k}^t = P(x_i|z_k^t) \tag{6}$$

### 4.3.3 Retrieval Model for Extended Topic-biased Transformation

Fig. 3 gives the demonstration of the retrieval process under the case of different topic distributions in both the spaces. During the query time we utilize the topic transformation matrix $\mathbf{H}$ to convert $\mathbf{q^{Z^b}} = [P(z_1^b|q), ..., P(z_K^b|q)]$ to an expected topic distribution of the query $q$ in the target space (see the portion of Fig. 3 marked in red color), denoted as $\boldsymbol{\lambda_q'}$ (see Eq. 7). Then we use this transformed topic distribution as weights for combining (summing) the results transformed using different term transformation matrices (weight assignment is shown by black dashed lines in Fig. 3).

$$\begin{aligned} S_{sim}(\mathbf{q},\mathbf{v}) &= \sum_{k=1}^{K} \lambda_{q,k}' \cdot cos(\mathbf{M_k q}, \mathbf{v}) \\ \text{where } \lambda_{q,k}' &= \text{the } k\text{th element in } \boldsymbol{\lambda_q'} \\ \boldsymbol{\lambda_q'} = \mathbf{H q^{Z^b}} &\text{here } \mathbf{q^{Z^b}} = [P(z_1^b|q), ..., P(z_K^b|q)] \end{aligned} \tag{7}$$

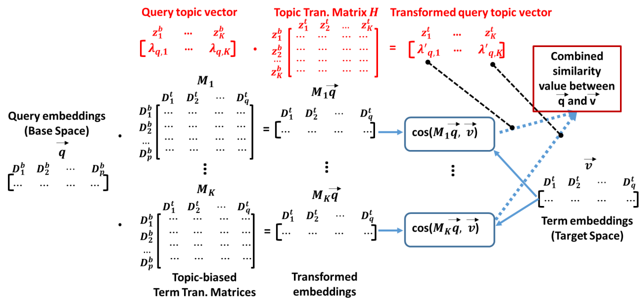


**Figure 3: Computing similarity between query $q$ and term $v$ in the target space using the topic-biased term transformation plus the topic transformation.**

# 5. EXPERIMENTAL SETUP

## 5.1 Datasets and Data Segmentation

For the experiments we use Wikipedia due to its large number of described entities and concepts, and due to its reasonably objective writing style. We downloaded the latest dump of English Wikipedia as of September 1st, 2016[4]. It contains more than 5 million articles in XML format with the decompressed size being approximately 40GB. To test the performance of searching across geographical areas, we segment the dataset by countries. In particular we extract the data specific to the tested countries using the category information according to the following procedure. (1) We first select all the categories and subcategories which contain the name or variant name of each country. For example, for collecting data for Japan, we extract all the categories containing the keyword "Japan' (e.g., Prime Ministers of Japan, World Heritage Sites in Japan, etc.). (2) Then we extract all the articles under these categories to form the dataset for the country. (3) Finally, we remove duplicates. Also, before the experiments, we discard the pages which are shared by the datasets of the two countries being tested. In the current experiments, we test the case of searching across Japan and United States. This choice is motivated by the fact that there are significant cultural, social and geographical differences between the both countries. Also, since the authors of this paper are based in Japan it is relatively easy to recruit subjects with good knowledge about Japan and reasonably good knowledge about USA. After collecting the data of Japan and that of United States in the way described above, we train the vector spaces for Japan and USA separately using Word2Vec[5] (described in Sec. 3.1 and schematically portrayed in Fig.4). The dataset statistics are shown in Tab. 1.

**Table 1: Statistics of Data Sets**

| Country | #Categories | #Articles | #Vocab. |
|---|---|---|---|
| Japan | 7,518 | 216,165 | 221,177 |
| USA | 18,212 | 1,017,327 | 563,675 |

**SFC Extraction.** As discussed in Sec. 3.2, *Shared Frequent Concepts* are extracted and used as anchors to map the two vector spaces. In the experiments, we propose to extract hypernyms separately from Japan related corpus and the USA related corpus, and then to take their overlap (as the terms must be shared) as the SFCs. We extract hypernyms from the first paragraph[6] of each article by applying syntactic pattern matching[7]. Hypernyms in the lead paragraphs are likely to denote general concepts.

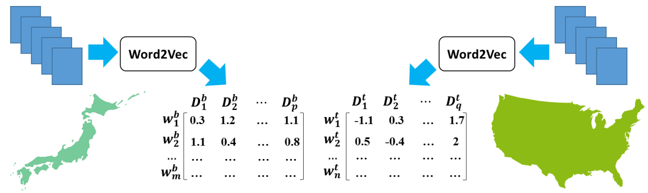


**Figure 4: Training word embeddings on separate datasets related to different countries.**

## 5.2 Test Sets

As far as we know there is no standard test bench for spatial counterpart detection. We manually create the test sets containing queries and then apply pooling technique. As we focus on the task of searching from Japan to USA, the queries are then representative for Japan. The selection process is mainly based on the criteria that (1) the query should exist in the vocabulary of Japan's dataset; (2) the query should be specific and common in Japan (trivial terms such as `car` or `water` which are global in the sense of being ubiquitous and semantically equivalent in any country will not be tested); (3) for the ease of evaluation, we try to select queries which might have correspondence in USA, otherwise, it is hard to evaluate the answers. Based on these criteria, we create test sets composed of 100 queries[8] covering four types of terms: *persons*, *locations*, *objects* and *non-entities*. We select queries across diverse topics to test methods in relation to wide range of possible topics and senses (e.g., politics, business, sports, local cultures, etc.). As for the location type queries, we try to cover different granularities and subtypes of location spots (particular cities, rivers, mountains, etc.). Aggregated statistics and query examples are illustrated in Tab. 2.

We have evaluated the total of 10, 103 answers for the experiments. In particular, we have leveraged the pooling technique [18] by pulling the top 20 results from 8 different systems (proposed methods and baselines as listed in Sec. 5.3.2). Three annotators judged every result in the pool as for whether it is similar/correspondent to the queried entity, producing, in total, 30, 309 judgments. The annotators did not know which systems generated which answers as all the terms from the pool were ordered alphabetically for each query. They were allowed to utilize any external resources or use search engines in order to verify the correctness of the results.

Each annotator took on average 70 hours[9] for completing the annotation task due to the need for learning about each

[4]https://dumps.wikimedia.org/enwiki/20160901/

[5]To train the term embedding models, we use bigrams, 200 dimensions, minimum frequency of terms equal to 5 and the window length of 5 for both datasets.

[6]Lead paragraphs of articles usually contain definitions or summaries of entities in Wikipedia.

[7]We focus on "is-a" relation including its variants, such as "was a", "are", "were", "known as', "regarded as", "called" etc.

[8]Available at: http://tinyurl.com/jydgat3

[9]On average, over 40min was needed to evaluate the pooled results for one query.

entity and searching for related details. A term was evaluated by giving score from 0 (worst) to 5 (best) according to the extent of similarity and correspondence between query and an answer term[10]. The final score of a term in pool is determined by taking the average score of the annotators. The average Fleiss' Kappa [7] is 0.79, indicating *substantial agreement* across the raters (values above 0.61 are considered as substantial agreement [13]).

**Table 2: Statistics and Examples of Queries**

| Type | #Q | Pool size | Examples |
|---|---|---|---|
| Person | 23 | 2,342 | Shinzo Abe, Hayao Miyazaki |
| Location | 25 | 2,761 | Okinawa, Ginza, Mount Fuji |
| Object | 26 | 2,493 | Toyota, JAXA, Shinkansen |
| Non-entity | 26 | 2,507 | tofu, wasabi, kimono, manga |

## 5.3 Evaluation Measures and Tested Methods

### 5.3.1 Evaluation Measures

We use Discounted Cumulative Gain (DCG)[11] at the following ranks: @1, @5, @10, @15 and @20. Moreover, we also compare all the methods using Mean Reciprocal Rank (MRR). The reciprocal rank of a query's response is the multiplicative inverse of the rank of the first correct answer being the highest ranked result whose score is equal or above 4.

### 5.3.2 Tested Methods

**Baselines.** We set up three baselines:

(1) **Word2Vec without transformation (W2V-J).** This method uses distributional representation for representing word semantics, same as the proposed methods do. However, it does not use any transformation. In particular, this method combines the data from the base dataset and the target dataset to obtain one vector space based on the joint dataset. The similarities between query and terms in the target space are computed in the joint term embedding model. Testing this method in comparison to the proposed methods will allow to answer the question about the necessity of the transformation.

(2) **LDA without transformation (LDA-J).** This baseline is similar to **W2V-Joint**. The difference is that it uses LDA for dimensionality reduction[12]. Unlike in the Neural Network approach (Word2Vec), each dimension of LDA-based word embedding corresponds to a given topic and the values of the embeddings represent the weights of the term belonging to each topic.

(3) **Direct Transformation (DT).** This method first locates the query in the base vector space by computing the distance from the query to the anchor points (SFCs discussed in Sec. 3.2), and then detects the counterparts in the target space by finding the ones which have the most similar distances from the anchor points to the query. It can be understood as a simple version of term transformation. Yet, instead of using the transformation matrix, it directly computes relative positions of query to the anchors.

[10]The answers that are clearly not related to the target geographical area such as cities located in other countries than the target country received score equal to 0.

[11]DCG is used instead of nDCG since the evaluated result lists are of the same lengths and the ideal ordering of results is unavailable.

[12]We use 200 as the number of topics and we remove terms with frequency less than 5.

**Table 3: Main Results in DCG@1,5,10,15,20 and MRR. Our methods significantly (p<0.01) outperform all the baselines across all the metrics. † indicates statistically significantly better performance than GT (p<0.1) (‡ represents the case of p<0.05). Results for methods -C which are marked with * are statistically significantly (p<0.05) better than the ones for variant methods -D.**

| Method | @1 | @5 | @10 | @15 | @20 | MRR |
|---|---|---|---|---|---|---|
| W2V-J | 0.61 | 2.33 | 3.52 | 4.66 | 5.55 | 0.218 |
| LDA-J | 0.28 | 0.66 | 1.00 | 1.23 | 1.43 | 0.072 |
| DT | 0.49 | 1.29 | 1.84 | 2.18 | 2.40 | 0.137 |
| GT | 1.84 | 5.54 | 7.92 | 9.57 | 11.15 | 0.416 |
| TT-C | 1.92* | **6.16***† | **8.49***† | **10.52***‡ | **12.04***‡ | 0.442*‡ |
| TT-D | 1.59 | 5.61 | 7.79 | 9.43 | 10.84 | 0.414 |
| ETT-C | **1.93*** | 5.81* | 8.25* | 10.14*† | 12.01*‡ | **0.446***‡ |
| ETT-D | 1.30 | 5.23 | 7.22 | 8.78 | 10.43 | 0.350 |

**Proposed Methods.** We test 5 proposed methods.

(1) **General Term Transformation** (**GT**) bridges the two vector spaces, however without biasing the transformation to different topics (see Sec. 3).

(2) **Topic-biased Transformation** (**TT**) trains transformation matrices separately for different topics under the assumption that the topic distributions of spatial counterparts remain the same. We test two variants of this method: (i) one that uses the weighted results from all the topics (Eqs. 3 and 4) (**TT-C**) and (ii) one that considers only the results from the most dominant topic of the query (**TT-D**). The latter is computed also by Eqs. 3 and 4. However in Eq. 4, we use only the results from the dominant topic instead of the weighted sum of results over all topics.

(3) **Extended Topic-biased Transformation** (**ETT**) considers also topic transformation but without the assumption of same topic distributions across spaces, hence, it is the closest to realistic scenarios. Similar to the above, we also test two variants for this method: (i) the one with the weighted combination from the results of all the topics (Eqs. 5, 6 and 7) (**ETT-C**) and (ii) one using solely the results from the most dominant topic (**ETT-D**). The latter is calculated in the similar way as **TT-D**.

# 6. EXPERIMENTAL RESULTS

Tab. 3 shows the results of search from Japan to USA by all the tested methods. We also display examples for each query type in Tab. 7. The main observation is that all our proposed methods statistically significantly (p<0.01) outperform the baselines and that **ETT-C** performs best according to DCG@1 and MRR, while **TT-C** outperforms the other methods for DCG@5, @10, @15 and @20. We discuss the results in detail below.

## 6.1 Necessity of Transformation

As mentioned before **W2V-J** and **LDA-J** do not perform any transformation. Although **W2V-J** performs better than the other two baselines (**LDA-J** and **DT**), it attains only about half of the DCG score at different ranks when compared to the proposed methods. Since **W2V-J** and **LDA-J** train the word representation models by combing the datasets of the two countries, they mix concepts (or topics) in the two spaces without considering their specific aspects in each country. In such sense, **W2V-J** loses the information of relative positions of terms within each semantic space, and **LDA-J** ignores the topic shift between two corpora.

## 6.2 Benefit of Transformation Matrix

By analyzing the results of **DT**, we observe that the direct mapping by measuring the relative distance to the anchors in the two vector spaces can hardly achieve desired results. Unlike the proposed method **GT**, **DT** performs term mapping without any learning process (see Sec. 3.2), simply based on Euclidean distance of terms to the anchors. The worse results indicate that the relationship between anchors (pivot points) and other terms (non-pivot points) is nested in the vector spaces and thus more complex. Therefore, it is necessary to have a good learning process to obtain a correct mapping function (e.g., transformation matrix).

## 6.3 Importance of Topic-biasing

Comparing the results of the proposed methods **TT-C** vs. **GT** and **ETT-C** vs. **GT**, we found that **TT-C** outperforms **GT** by 4% (@1), 11% (@5), 7% (@10), 10% (@15) and 8% (@20). **ETT-C** is better than **GT** by 4.5% (@1), 5% (@5), 4% (@10), 6% (@15) and 7.7% (@20). These findings support our motivation behind introducing the topic-biased term transformation, namely, a general matrix is not enough for transforming terms. This is because Eq. 1 performs a global optimization scarifying precision in each specific case. Through topic-biasing, every individual matrix offers better precision in mapping terms belonging to a given topic than the general transformation matrix does in **GT**.

We have also found that merging results from all the topics is better than relying only on a dominant topic as evidenced by better performance of **TT-C** over **TT-D** and **ETT-C** over **ETT-D** across all the metrics (stat. signif. at $p<0.05$).

## 6.4 Scenarios of Using Topic Transformation

According to Tab. 3, it seems that **ETT** cannot enhance the performance over all the queries. We explore now situations when the topic transformation (or **ETT**) helps to improve the results. In particular, we analyze the results from the viewpoint of query frequency and topic entropy.

**Frequency based Evaluation.** To analyze the influence of frequency, we divide queries into three equal size groups by their counts in the base dataset (i.e., Japan corpora): high frequency, medium and low frequency. We regard queries with high frequency as “easy” cases, while relatively rare queries (low frequency) as “hard” cases. We then recalculate the DCG@20 scores for the queries of each group as illustrated in Tab. 4. We can observe that our method **TT-C** achieves the best results under easy cases. However for the medium and hard cases, the extended methods considering topic transformation perform best (**ETT-D** in the medium cases and **ETT-C** in the hard cases). We can then conclude that when a user searches for counterparts of low frequency queries, selecting **ETT** is the best option as it is most likely to return the best answers.

**Topic Entropy based Evaluation.** Besides the frequency, we also examine the impact of the skewness of topic distributions in queries. The topic entropy of each query is computed by $\sum_K(-P(z_i^b|q) \cdot \log_2(P(z_i^b|q))$. A query characterized by a low entropy (easy case) has skewed topic distribution, which means that the query represents rather topically clear information. On the other hand, a high entropy query (hard case) whose topic distribution approaches uniform distribution (the query has similar assignment to multiple topics) relates to topically ambiguous information. According to the DCG@20 scores in each group (see Tab. 5), we obtain similar conclusions as before that our extended methods **ETT-D** and **ETT-C** perform best in the hard cases and medium cases. As for the easy case, **TT-C** achieves the highest performance same as in Tab. 4.

**Table 4: DCG@20 for Easy, Medium and Hard Cases by Frequency**

| Method | High Freq. (Easy Cases) | Medium | Low Freq. (Hard Cases) |
|---|---|---|---|
| W2V-J | 7.97 | 2.47 | 6.21 |
| LDA-J | 2.91 | 1.19 | 0.17 |
| DT | 2.86 | 1.70 | 2.64 |
| GT | 14.53 | 6.36 | 12.55 |
| TT-C | **15.37** | 7.55 | 13.19 |
| TT-D | 14.08 | 7.10 | 11.33 |
| ETT-C | 15.02 | 7.56 | **13.44** |
| ETT-D | 10.11 | **8.27** | 12.92 |

**Table 5: DCG@20 for Easy, Medium and Hard Cases by Topic Entropy**

| Method | Low Entropy (Easy Cases) | Medium | High Entropy (Hard Cases) |
|---|---|---|---|
| W2V-J | 6.27 | 5.26 | 5.11 |
| LDA-J | 2.20 | 1.00 | 1.08 |
| DT | 1.74 | 3.03 | 2.43 |
| GT | 12.37 | 10.94 | 10.12 |
| TT-C | **13.74** | 10.53 | 11.84 |
| TT-D | 12.85 | 10.11 | 9.55 |
| ETT-C | 12.56 | **11.18** | 12.28 |
| ETT-D | 9.34 | 9.59 | **12.37** |

## 6.5 Evaluation of Query Types

We also evaluate the methods from the perspective of the query types. Tab. 6 demonstrates that our proposed method **TT-C** has the better performance in searching for spatial counterparts in the category of persons, objects and non-entities. However, for the locations, **ETT-D** performs best. It might be due to locations (e.g., cities) being characterized by relatively ambiguous topics, that is, locations tend to be mentioned in a variety of diverse topics and in many different contexts[13]. Hence, the topic transformation becomes useful in the case of such queries. Interestingly, we have also found that, according to DCG@20 scores for different query types, non-entity queries are easiest resulting in the highest scores for all the compared methods, which might be attributed to the relatively clear meaning of non-entities. On the other hand, locations seem to represent the most difficult task for all the methods. The reason might be again the relatively higher topical ambiguity of locations as mentioned above.

**Table 6: Evaluation over Query Types (DCG@20)**

| Method | Person | Location | Object | Non-entity |
|---|---|---|---|---|
| W2V-J | 8.05 | 1.23 | 4.01 | 8.68 |
| LDA-J | 0.12 | 0.24 | 1.92 | 3.24 |
| DT | 3.09 | 1.89 | 1.05 | 3.45 |
| GT | 14.92 | 5.28 | 8.99 | 15.14 |
| TT-C | **15.26** | 6.62 | **10.63** | **15.43** |
| TT-D | 11.81 | 6.60 | 10.40 | 14.27 |
| ETT-C | 15.00 | 7.79 | 10.13 | 14.91 |
| ETT-D | 14.68 | **9.14** | 8.11 | 9.93 |

**Case Studies.** Looking at examples shown in Tab. 7, our methods reveal better performance in detecting good

[13] 88% of location queries are in Medium and High Entropy categories.

counterparts at top ranks, e.g., **TT-C** and **ETT-C** can detect `NASA` as counterpart of `JAXA` at the first rank. Another interesting example to mention is `tofu` (as in this paper's title). When researching about its counterparts, we found evidences outside of Wikipedia that support `cheese` as its counterpart due to the similar preparation method, consistency and role in the local cuisine[14]. Apart from these examples, more implicit across-countries correspondences can be found such as `Akira Kurosawa`, who is regarded as the most important and influential Japanese film director. The proposed methods **TT-C** and **ETT-C** returned `William Wyler` as his counterpart in USA. Although, the two persons are not co-appearing in their Wikipedia's articles, they have many resemblances including the fact that Akira Kurosawa has been influenced by William Wyler[15].

## 7. DISCUSSIONS

In the current implementation we focus on English language, and, by this, we avoid the **problem of language translation** such as translating from Japanese to English. More complex methods could be however designed to first translate from the base language to the target one and then to perform the mapping as described in this paper.

Another issue relates to queries for which there are no good counterparts (e.g., highly unique entities very specific to one country). While our methods will still return some results in such cases, the associated similarity scores should be interpreted by searchers as **confidence levels of results**.

An interesting research problem is to provide **explanations for annotating and understanding the returned results**. Perhaps, the simplest way to implement it could be by selecting terms that are frequently co-occurring with a given returned spatial counterpart and that also belong to query's dominant topic. Alternatively, algorithm similar to the one introduced in [23] could be applied.

The proposed methods can be extended in several different directions. First, it could be possible to design search algorithms for **finding corresponding relationships in different geographical spaces**. The input could be then in the form of term set (e.g., {city,city}, {building,city}, {food,country}).

Second, our methods could be extended to enable **query-by-region search mode**. Namely, users could point on a map or select some area within one country and the system would then automatically suggest similar point or area in another country. This could be realized by collecting geo-entities in the selected region as well as any strongly related with it concepts and then searching for regions in the target country with similar geo-entities and concepts.

Lastly, the topic-based transformation could be utilized to enable **topically-focused analogical search**. This kind of search could be useful when users wish to find counterparts under given viewpoint (e.g., cities with similar industry).

## 8. RELATED WORK

Several researchers [5, 12, 11, 16] have approached domain adaptation task. Blitzer *et al.* [5] proposed a Structural Correspondence Learning (SCL) to identify correspondences among features from different domains by modeling their correlations with pivot features. The method was proved to perform well in a discriminative framework, such as in the task of PoS-tagging. Similarly, Kato *et al.* [12, 11] proposed to utilize Relative Aggregation Point (RAP) such as average price, maximum/minimum cost, restaurant categories etc. in different domains as features to detect a corresponding restaurant in another city. Both of these approaches were done in a discriminative learning manner where a conditional probability of the instances in a domain was estimated and classified into a certain class. However, these approaches only work for the data where the instances are already classified or the distributions of the instances over categories are known in a domain. For many datasets, such as news archives, online reviews, encyclopedias where the entities are unstructured or the entities are not represented by any fixed attributes, one needs to leverage other information to solve the domain adaptation problem. Unlike these researches, we propose a general framework by only leveraging the semantics of terms and their relative positions in each semantic space to perform transformation. Our methods can be applied to any orthogonal raw-text datasets while the query can be any term (e.g., city, person, object, culture).

The problem of geographic analogues have been approached by several researches. A recent example, Frankenplace [1], established an interactive interface for exploratory search in which words are mapped to a grid map by utilizing the co-occurrence between words and the place names in documents. In such sense, the "similarity" between words is determined by whether they occur in the same documents. Such an approach shows however promising results in searching for geographical counterparts only if the document which contains the candidate counterpart refers also to the query. Although some queries could be mentioned with their counterparts in the same Wikipedia article (e.g., there is a link from Wikipedia article about "Yomiuri Giants" to the one about "New York Yankees") most of the time this is not the case. When it comes to the cases where query and counterparts do not co-occur or co-occur very rarely, their approach does not work as it does not consider the contexts or semantics of each word. Compared to [1], our method does not only consider the semantics of each word captured by its context but it also relaxes the limitations of term co-occurrence. Moreover, our approach can work for unstructured text, such as news articles without the need to analyze any links among documents.

Another related line of work dealt with the task of finding similar terms across time [2, 10, 9, 21, 22]. For example, Berberich *et al.* [2] introduced Hidden Markov Model based approach to detect two similar terms at different time periods. Kalurachchi *et al.* [9] discovered semantically equivalent concepts by applying association rule mining and assuming that concepts referred by similar verbs are related. Other works compared temporal snapshots of Wikipedia for detecting changes in terms over time [10]. Zhang et al. [21, 22] used neural network based term representations and established transformation matrix for finding semantic counterparts of present entities. Our objective is however different as we attempt to detect semantically corresponding objects across different spaces.

Research on analogical relation detection [6, 20, 19] has also some relation to this work. Structure Mapping Engine (SME) [6] - the original implementation of the well-known Structure Mapping Theory (SMT) [8] that explains

[14] http://www.chinaexpat.com/2008/06/04/tofu-vs-cheese.html/ http://www.japan-guide.com/e/e2045_tofu.html
[15] http://www.nextpix.com/v1_1/salon/kurosawa.html

**Table 7: Top 10 terms by each method for selected queries. Terms in bold font are judged as correct (average score⩾4) by annotators.**

| W2V-J | LDA-J | DT | T | TT-C | ETT-C |
|---|---|---|---|---|---|
| | | | **Haruki Murakami** | | |
| Eiji Yoshikawa | loire | mystery writer | **Thomas Pynchon** | **Thomas Pynchon** | **Thomas Pynchon** |
| Yasunari Kawabata | Pasadena | dramatist | Andr Gide | **Willa Cather** | **John Updike** |
| Rynosuke Akutagawa | Stephen Elliott | ethnomusicologist | **Raymond Carver** | **John Updike** | **Willa Cather** |
| Donald Richie | game show | feminist writer | **Vladimir Nabokov** | **William Styron** | **Raymond Carver** |
| **Michael Crichton** | Gregory Gerrer | Leo Tolstoy | Saul Bellow | mystery writer | Leo Tolstoy |
| Banana Yoshimoto | thurber | folklorist | **Theodore Dreiser** | **William Faulkner** | **William Styron** |
| Rytar Shiba | Filip Bobek | **John Updike** | **Willa Cather** | **Vladimir Nabokov** | **William Faulkner** |
| Alan Moore | resor | cultural critic | anthologist | **Raymond Carver** | **Theodore Dreiser** |
| **Raymond Carver** | Epileptic | Stephen Schwartz | **Randall Jarrell** | Thomas Hardy | Harold Bloom |
| historical fiction | Hinamori | writer poet | **Don Delillo** | John Ruskin | **Ralph Ellison** |
| | | | **Mount Fuji** | | |
| Mount Miwa | Keisei | Palouse Hills | sand dune | happy isles | happy isles |
| mount haku | Yoshioka | Halcott | **Mount Evans** | **Mount Evans** | Creek Canyon |
| Konohanasakuya Hime | prefabricated | **Steens Mountain** | **Steens Mountain** | Ventura Boulevard | **Mount Evans** |
| Fushimi Kyoto | Tregurtha | Lake Pillsbury | happy isles | **Santiam Pass** | **Mount Shasta** |
| hakone | wildcat | Garibaldi Lake | Halcott | kingshighway | Foothill Boulevard |
| suijin | interpose | Clarno | roundtop | Glenwood Canyon | sugarloaf |
| iide | waldron | southwestern edge | toroweap | Sutro Tower | roundtop |
| Nagara River | dortch | southwestern | Suksdorf Ridge | Foothill Boulevard | Sutro Tower |
| Niigata Prefecture | Banpaku | Alvord Desert | outcropping | Mallory Square | **Tioga Pass** |
| Hiruzen | qcc | **Olallie Butte** | hellroaring | **Tioga Pass** | **Steens Mountain** |
| | | | **JAXA** | | |
| **Spacex** | mach2 | Ffrdc | technology | **NASA** | **NASA** |
| ISAS | hasp | ns ep | **NASA** | technology | technology |
| isro | transponder | information dissemination | microgravity | Lincoln Laboratory | microgravity |
| **NASA** | stairstep | environmental monitoring | ligo | space telescope | rocket propulsion |
| diwata 1 | alcms | supportability | rocket propulsion | manned spacecraft | grid computing |
| space probe | osta | communication networks | remote sensing | ligo | advanced computing |
| voyager 2 | spacelab | metrology | oco | manned orbital | afams |
| ISS | aerospike | systems integration | lunar exploration | heliophysics | spawar |
| human exploration | uuv | business continuity | crystal growth | oco | remote sensing |
| blue origin | suvorov | wireless communications | heliophysics | NASA manned | midcourse |
| | | | **tofu** | | |
| soy sauce | restaurateurs | **swiss cheese** | vegetable | **cheese** | country ham |
| deep fried | fertilize | cole slaw | celery | vegetable | celery |
| broth | tonkotsu | rice flour | olive oil | sausage | cole slaw |
| pork | heirloom | corn masa | biscuits | biscuits | corn husk |
| fermented | meals | provolone | corn masa | meat | vegetable |
| sweet potato | chilled | country ham | cole slaw | pork | **cottage cheese** |
| grilled | broiled | basic ingredients | garlic | bread | olive oil |
| tonkatsu | watermelon | blue corn | **cottage cheese** | olive oil | whole wheat |
| aburage | panisse | paper towels | potatoes | salad | pumpkin pie |
| green tea | bourdain | garnishes | mayonnaise | potatoes | sauerkraut |

how humans tend to reason using analogy - is perhaps one of the oldest examples of computational approaches to analogy. Subsequently, Turney proposed Latent Relational Mapping Engine (LRME) [20] that extracts lexical patterns in which words tend to co-occur in order to measure relational similarity. The difference of these approaches and our methods is that the former are always based on a single document dataset. In such settings, contextual information specific to particular country is lost, as we also demonstrate in our experiments. Finally, several types of models have been proposed for the task of proportional analogy detection [20]. However, the objective in this case is to extract an object that can fit into equation $a : b :: c : d$ where one of the four constituents is unknown.

# 9. CONCLUSIONS

Nowadays, users often search for information related to distant and unknown places. To decrease the problem stemming from the vocabulary gap we propose query suggestion mechanism based on automatic transformation of concepts from one spatial area to another. The problem is not trivial due to diverse contexts of semantically similar terms within different spaces as demonstrated by poor performance of approaches relying on a joint dataset. We introduce several unsupervised methods for mapping terms from different places such as different countries. An important characteristics of our approach is that it works on raw text collections without the need for utilizing knowledge bases or any supervision.

In future we will experiment with other countries and document collections, e.g., news articles or tweets. We will also try to implement some of the extensions listed in Sec. 7.

# 10. ACKNOWLEDGEMENTS

This work was supported in part by Grants-in-Aid for Scientific Research (Nos. 15H01718, 15K12158) from MEXT of Japan and by JSPS KAKENHI (Grant-in-Aid for JSPS Fellows): “A Study on Temporal Data Mining and Object Search in Big Archival Data” (No. 16J03128).

# 11. REFERENCES

[1] B. Adams, G. McKenzie, and M. Gahegan. Frankenplace: interactive thematic mapping for ad hoc exploratory search. In *Proceedings of the 24th International Conference on World Wide Web*, pages 12–22. ACM, 2015.

[2] K. Berberich, S. J. Bedathur, M. Sozio, and G. Weikum. Bridging the terminology gap in web archive search. In Proc. of WebDB, 2009.

[3] D. M. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[5] J. Blitzer, R. McDonald, and F. Pereira. Domain adaptation with structural correspondence learning. In *Proc. of EMNLP*, pages 120–128, 2006.

[6] B. Falkenhainer, K. D. Forbus, and D. Gentner. The structure-mapping engine: Algorithm and examples. *Artificial intelligence*, 41(1):1–63, 1989.

[7] J. L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382, 1971.

[8] D. Gentner. Structure-mapping: A theoretical framework for analogy*. *Cognitive science*, 7(2):155–170, 1983.

[9] A. C. Kaluarachchi, A. S. Varde, S. Bedathur, G. Weikum, J. Peng, and A. Feldman. Incorporating terminology evolution for query translation in text retrieval with association rules. In Proc. of CIKM, pages 1789–1792, 2010.

[10] N. Kanhabua and K. Nørvåg. Exploiting time-based synonyms in searching document archives. In Proc. of JCDL, pages 79–88, 2010.

[11] M. P. Kato, H. Ohshima, S. Oyama, and K. Tanaka. Search as if you were in your home town: Geographic search by regional context and dynamic feature-space selection. In *Proc. of CIKM*, pages 1541–1544, 2010.

[12] M. P. Kato, H. Ohshima, and K. Tanaka. Content-based retrieval for heterogeneous domains: Domain adaptation by relative aggregation points. In *Proc. of SIGIR*, pages 811–820, 2012.

[13] R. J. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174, 1977.

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In *Proc. of ICLR Workshop*, 2013.

[15] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representation of phrases and their compositionality. In *Proc. of NIPS*, pages 3111–3119, 2013.

[16] S. Pan and Q. Yang. A survey on transfer learning. *IEEE TKDE*, 22(10):1345–1359, 2010.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. Technical report, California Univ, San Diego La Jolla Inst. For Cog-nitive Science, 1985.

[18] J. K. Sparck and V. C. Rijsbergen. *Report on the need for and provision of an ideal information retrieval test collection*. British Library Research and Development Report 5266, Computer Laboratory, University of Cambridge, 1975.

[19] P. D. Turney. Expressing implicit semantic relations without supervision. *CoRR*, 2006.

[20] P. D. Turney. The latent relation mapping engine: Algorithm and experiments. *Journal of Artificial Intelligence Research*, pages 615–655, 2008.

[21] Y. Zhang, A. Jatowt, S. S. Bhowmick, and K. Tanaka. Omnia mutantur, nihil interit: Connecting past with present by finding corresponding terms across time. In *Proc. of ACL*, pages 645–655, 2015.

[22] Y. Zhang, A. Jatowt, S. S. Bhowmick, and K. Tanaka. The past is not a foreign country: Detecting semantically similar terms across time. *IEEE Transactions on Knowledge and Data Engineering*, 28(10):2793–2807, 2016.

[23] Y. Zhang, A. Jatowt, and K. Tanaka. Towards understanding word embeddings: Automatically explaining similarity of terms. In *Proceedings of 2016 IEEE International Conference on Big Data*, pages 823–832, 2016.