

Seminar Component

Name of author : Abd. Hadi Asfarangga and Fitri Andri Astuti

Organization : BPS-Statistics Indonesia

Contact address : Jl. dr. Sutomo No. 6-8 Jakarta 10710, Indonesia

Contact phone : +6221 3841195/3810291-4

Email : hadi.asfarangga@bps.go.id and fitri.andri@bps.go.id

Title of Paper

Utilizing Google Places API in Enhancement of NPISH Directory

Abstract

NPISH as one of the institutional sectors engaging in the transaction and owning assets have important roles in the economy. NPISH contribute quite significantly and grow fast especially in Indonesia economy. Besides that, NPISH also concern in development of household wellbeing to reduce poverty since they produce mainly non market output which would become part of actual final consumption of household. Regarding these reasons, NPISH statistics is crucial and it needs to be more developed and improved. One of area which could be enhanced is directory of NPISH. It functions as basis for sampling survey and population of NPISH. BPS as national statistics office of Indonesia has been developing NPISH directory as we publish NPISH final consumption separately from household final consumption in GDP by expenditure. In constructing the directory, we employ particularly administrative data, i.e. ministry registrations either in national or regional level. Other sources actually exist such as village potential survey, but it only provides the number of NPISH. In this digital era, many data are available in internet such as google places API which could be used for development of NPISH directory. We extract data from google places API containing same database used by google maps to obtain list of potential NPISHs. Then, these potential NPISHs are clustered to distinguish whether they are real NPISHs or still ambiguous units. These ambiguous units need to be clarified and identified to determine treatment of every case. Finally, we run matching process with criteria of similarity rate to get list of NPISHs not included yet in current NPISHs directory. This result becomes new supporting data in NPISH directory and population in order to producing statistics better.

Keywords: NPISHs directory, Google Places API

I. Contents

I. Contents	2
II. Introduction	3
III. Directory Construction	3
A. Concept of NPISH	3
B. Data Extraction	4
1. Subheading A.1	4
2. Subheading A.2	4
3. Subheading A.3	4
C. Data Clustering	4
1. Subheading B.1	4
2. Subheading B.2	4
D. Data Matching	4
IV. Conclusion	4
V. References	4

II. Introduction

Demand on data and information rises rapidly nowadays since public societies become more literate in statistics in the era of openness information. National statistics office should move forward to meet the user needs particularly in term of data variety and completeness. BPS-Statistics Indonesia continuously strives developing official statistics in order to fulfill the public needs on data and information such as providing NPISH statistics which is in line with SNA 2008 recommendation.

Non-Profit Institutions Serving Households (NPISH) are one of the institutional sectors engaging in the transaction and owning assets in the economy. NPISH have several roles in the economy, i.e. production, consumption and investment actors. NPISH contribute quite significantly and grows fast especially in Indonesia economy. In 2017, NPISH had final consumption contribution at 1.18 percent to GDP with year-over-year growth of 6.91 percent. Besides that, NPISH also concern in development of household wellbeing to reduce poverty since they produce mainly non-market output which would become part of actual final consumption of household. Regarding these reasons, NPISH statistics is crucial and it needs to be more developed and improved.

One of area which could be enhanced is directory of NPISH. It functions as basis for sampling survey and population of NPISH. That directory can also be a supporting information for statistical business register. Statistics Indonesia has been developing NPISH directory as we publish NPISH final consumption separately from household final consumption in GDP by expenditure. In constructing the directory, we employ particularly administrative data, i.e. ministry registrations either in national or regional level. Other sources actually exist such as village potential survey, but it only provides the number of certain NPISH.

In this digital era, many data are available in internet that could be used for development of NPISH directory. Internet penetration rate of NPISH is assumed to be high enough as they mostly depend on donation or other transfer-in accounts in financing their activities. They need to make publicity to be notable by societies, government and corporations for gaining more donors and philanthropist.

One of powerful data in the internet is Google Place API containing the same data with Google Maps. It has over one billion active users of the system each year (Forbes, 2017). Hence, use of Google Place API is a good opportunity for helping national statistics office construct NPISH directory.

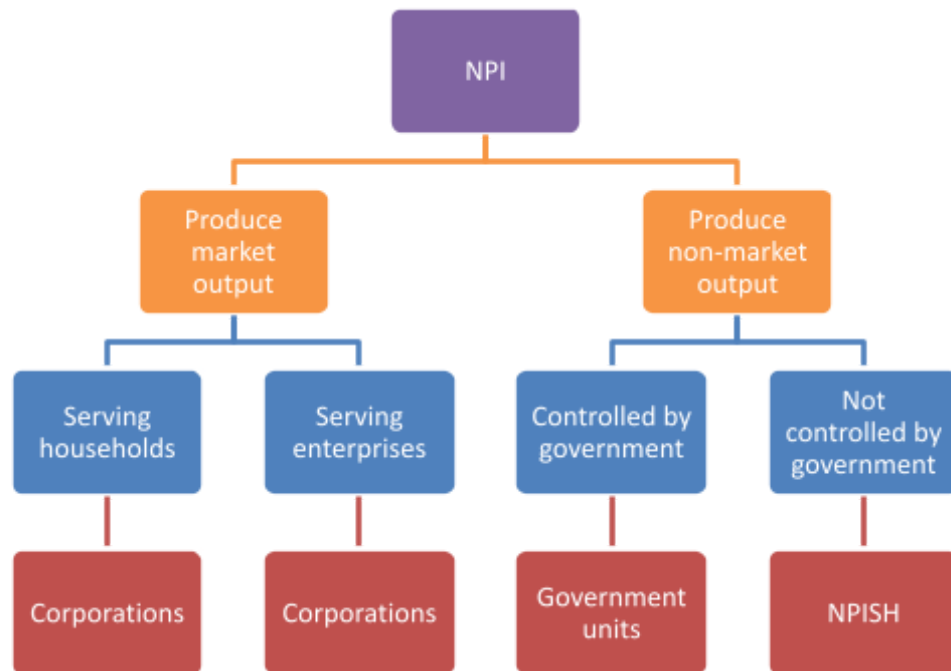
III. Directory Construction

A. Concept of NPISH

SNA 2008 divides institutional units into two main types, i.e. households and legal or social entities. Legal or social entities comprise corporations, government units and nonprofit institutions (NPI). SNA 2008 explains that NPIs are legal or social entities created for the purpose of producing goods and services but whose status does not permit them to be a source of income, profit or other financial gain for the units that establish, control or finance them. NPI generally have characteristics as follows (SNA 2008, page 72):

- Most NPI have existence recognized by law independently of the persons, corporations or government units that establish or control them. In some countries, NPI can be an informal entity whose existence is recognized by society but does not have any formal legal status.
- Many NPI are controlled by association whose members have equal rights.
- There are no shareholders with a claim on the profits of the NPI.
- The direction of an NPI is usually vested in a group of officers, executive committee or similar body elected by a simple majority vote of all the members.
- The members of the association controlling the NPI are not allowed to gain financially from its operations.

NPI can produce market and non-market output. The former is output sold at economically significant prices which are determined based on production cost and which can affect the demand of NPI products. In the meantime, the latter is output provided free or at prices that are not economically significant. NPI producing market output either serving households or serving enterprises constitute as corporations, while NPI producing non-market output can be government units or Non-profit Institutions Serving Households (NPISH). If NPI producing non-market output are controlled by the government, those NPI will be categorized as government, otherwise they will constitute as NPISH.



According to Classification of the Purposes of Non-Profit Institutions Serving Households (COPNI), NPISH can be grouped into nine activities elucidated as follows:

1. Housing
2. Health
3. Recreation and culture
4. Education
5. Social protection
6. Religion
7. Political parties, labor and professional organizations
8. Religion
9. Services n.e.c

In Indonesia, we also made classification of NPISH based on their types of institutions. The classification is described as follows:

1. Civil society organizations and political parties

Example: Muhammadiyah, Nahdlatul Ulama, Indonesia National Youth Committee (KNPI), Prosperous Justice Party (PKS), Gerindra Party, etc.

2. Social organizations

Example: orphanage, nursing home, rehabilitation center, etc.

3. Professional organizations

Example: Indonesian Doctor Association (IDI), Indonesian Dentist Association (IDGI), Indonesian Statistician Association (ISI), etc.

4. Recreational and cultural organizations

Example: supporters club, art club, etc.

5. Non-governmental organization

Example: Indonesian Forum for Environment (Walhi), Indonesian Consumers Foundation (YLKI), Institute for Research, Education and Economic and Social Information (LP3ES), etc.

6. Religious institutions

Example: mosque, church, temple, monastery, quran education training, etc.

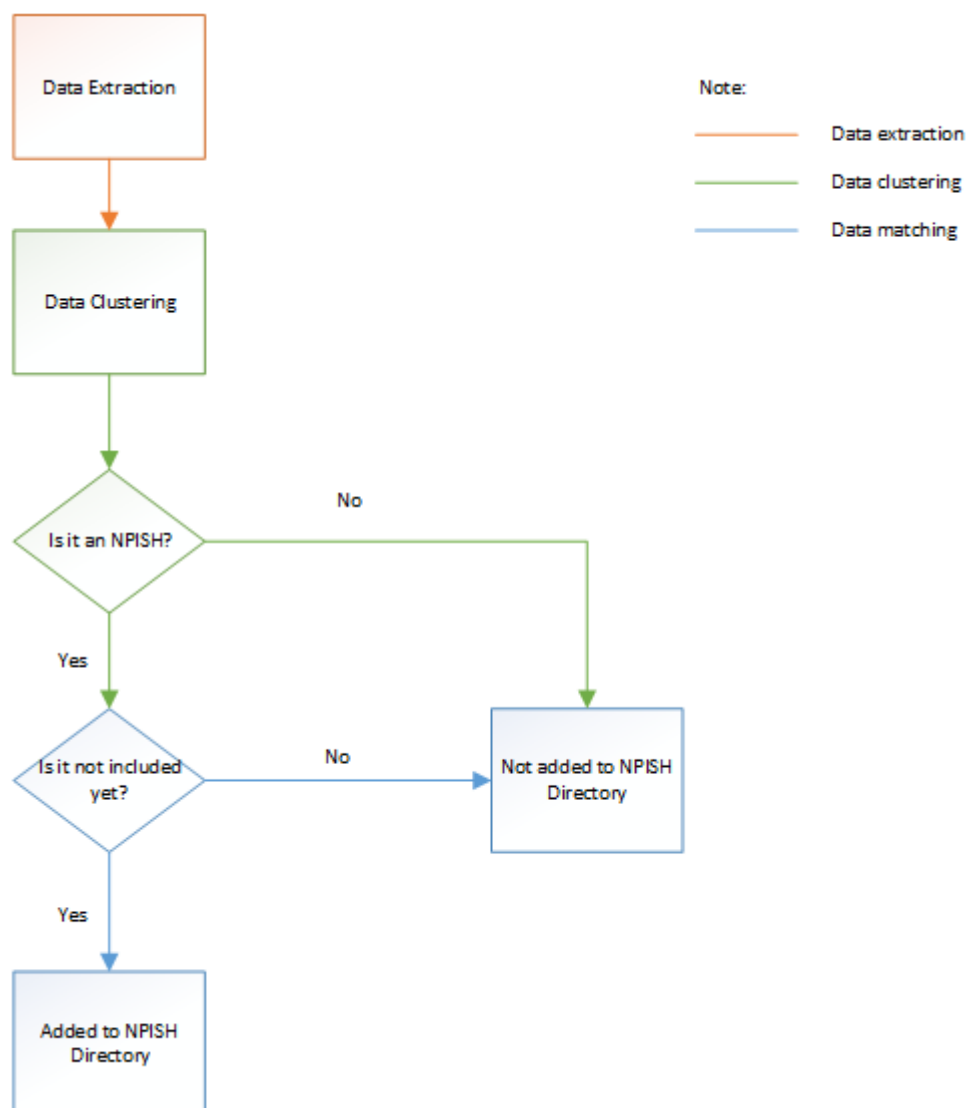
7. Humanitarian aid and scholarship organizations

Example: Indonesian Cancer Foundation Healthy Heart Foundation, Mercy Corps Indonesia, etc.

In this paper, we use the second classification for acquiring list of NPISH more easily. Indonesian NPISH usually have many activities with different COPNI classification. Nevertheless, we can still bridge their main activities to COPNI classification based on the purpose of NPISH establishment.

B. Data and Methodology

In constructing NPISH directory, we utilize Google Place API which is one of database owned by Google. We research and build NPISH directory in DKI Jakarta as prototype. Nonetheless, it can be duplicated for the entire provinces. We apply three steps of processing, i.e. data extraction, data clustering and data matching to obtain NPISH directory. The flow chart of those steps are depicted as follows:



1. Data Extraction

Firstly, we extract data from Google Place API which is accessed from our system. We developed a tool to search and retrieve NPISH data from Google Place API. This tool facilitate us gaining the data automatically without visiting the website.

In searching the NPISH, we enter several keywords associated with the kind of NPISH. The list of keywords is presented in the table below:

No.	Types of Institution	Keyword
1	Civil society organizations and political parties	Partai Politik

2	Social organizations	Panti Asuhan
3	Professional organizations	IDI
4	Recreational and cultural organizations	Ikatan Motor Indonesia
5	Non-governmental organization	LP3ES
6	Religious institutions	Masjid
7	Humanitarian aid and scholarship organizations	Yayasan Jantung Sehat

Tambahan: Yayasan

We use only seven keywords in this paper due to practical consideration, but we actually can expand the keywords to incorporate more potential NPISH.

After entering the keywords, the tool will retrieve all places or institutions related with them. Then, we obtain these lists of potential institutions saved in our database. This database comprise 8 variables explained as follows:

No	Field	Description
1	Id	Primary Key of the table in database
2	Id Prop	Identity of Province in Indonesia
3	Id Kab Kota	Identity of Regency/City in Indonesia
4	Name	Name of Institution
5	Type	Type of the address component
6	Address	Human-readable address of this place
7	Latitude	Geocoded latitude value for this place
8	Longitude	Geocoded longitude value for this place

We usually retrieve the data several times because Google restricts us to download only 1,000 free requests of institution search per 24 hour period.

2. Data Clustering

After extracting the data, the potential institutions then need to be clustered. This clustering process aims at classifying the potential institutions into group of NPISH and group of non NPISH. To cluster the potential institutions we employ supervised machine learning algorithm by using PHP library, i.e. uClassify. uClassify is a free text classification and can be accessed on www.uclassify.com. Besides that, this tool is also easy-to-use machine learning software even for lay user in computer programming because uClassify has friendly interface operated by click on it.

The PHP library is employed to submit a request of clustering the potential institutions in the uClassify. Before it can group them into NPISH and non NPISH, we will require some data training. Data training involves creating all wanted classes and training each class on the potential institutions constituting their coverage. In training data, we use two classes, ie. npish class containing the names of NPISH institutions and non_npish class containing the name of NPISH institutions. The name of NPISH and non NPISH institutions are obtained by taking sample of potential institutions then grouping them into NPISH or non NPISH manually by utilization of information on the internet.

We input some institutions in data training and the rest of institutions are predicted by uClassify. An institution will be group to certain class (NPISH or non NPISH) if that class has percentage more than 50 percent. The prediction results are examined by comparing them to manual classifying whether every institution is successfully predicted. Then, we calculate accuracy rate of the prediction by dividing the number of successful case to the total of tested institutions.

The data training shown in table 1.


Table 1


NPISH	NON NPISH
Yayasan Yatim Piatu Rasulullah SAW	Yayasan Pendidikan Saint John
Muslim Orphanage	YPAB - Yayasan Pemimpin Anak Bangsa
Indonesia Care for Cancer Kids Foundation	Yayasan Bhakti Abadi
Yayasan Hope Indonesia	Dana Pensiun Pertamina

Yayasan Al Kahfi Cabang Jakarta Pusat	Yayasan Yarsi
Yayasan Perkantas	Yayasan Kartika Eka Paksi
Yayasan Kanker Indonesia	Yayasan "Santa Theresia"
Yayasan Griya Asih	Yayasan Perguruan Cikini
Yayasan Masjid Meranti	Yayasan Bethel
Gedung Yayasan Kasih Bersaudara	Yayasan Administrasi Indonesia
Yayasan Nurul Iman Jafariah	Yayasan Rukun Istri Sejahtera Panti Asuhan Putra Setia
Yayasan Sahabat Anak	Yayasan PGI Cikini
Yayasan Panti Wanita Trisula Perwari	Dino Tour & Travel (Pulau Harapan)
Yayasan RCTI Jalinan Kasih	Yayasan Baby Sitter
Yayasan At-Taqwa	Yayasan BPK Penabur
Yayasan Kasih Bersaudara	Kantor Pengurus Pusat Yayasan Hang Tuah
Gurudwara Pasar Baru	YAYASAN STELLA MARIS
Indonesian Islamic Center Foundation	
Yayasan Masjid Istiqlal	
Yayasan At-Taufiq	
Yayasan Waqaf Al- Mujahirien	
Yayasan Baitul Halim	
YAYASAN MAKAM HABIB ALI	
Pelita Ilmu Foundation	
Yayasan Pelita Ilmu	
Yayasan Masjid An Nur	
Yayasan AIDS Indonesia	
Jakarta Indonesian Cancer Foundation	

Yayasan Peduli Tuna Daksa (Limb for The Limbless Center)	
Yayasan Sikh Temple Tanjung Priuk	
Light Dharma Foundation (Fo Kuang Ming Thang)	
Yayasan Seni Kehidupan	

uClassify Classifiers Translate Docs Pricing About


 cenolcantik Account Api keys Classifiers Profile

 npish Classify Classes Settings

Classify Text Classify Uri

Classify Text

Yayasan Amal Mulia Indonesia

 **Success** Show REST XML URL

npish
79%

not_npish
21%

3. Data Matching

Having been ensured that they are NPISH, the potential institutions will be matched with the current directory. This step is conducted to determine whether they will be added in the NPISH directory

In BPS, we acquire the current directory from the administrative data either in regional or national level. In this study, we use NPISH directory in DKI Jakarta, the capital of Indonesia.

We match the results of Google Place API whose the institutions have been clustered to the current directory by using fuzzy algorithm. The matching is robust to a wide variety of errors including spelling mistakes, abbreviations, synonyms and added/missing data. We add score of similarity by name and by address using tools of excel called add ins fuzzy lookup. We specify the threshold of similarity rate at xxx percent. If an institution has similarity rate less than xxx percent, it will be added in the NPISH directory and vice versa.

C. Result and Analysis

By using extraction tool which we have developed, list of potential institutions are obtained and saved in our database. An example of that list is presented below:

Data Clustering

Name	uji_npish	true/false
Yayasan Kesejahteraan Pendidikan dan Perumahan	0	1
Yayasan Dana Kami	0	1
Indonesian Islamic Center Foundation	1	0
Yayasan Bethel	1	0
Yayasan Fatmawati	1	0
Yayasan Al Istiqomah Tenggulun	1	1
Balai Pertemuan Yayasan Daarul Aitam	1	1
Yayasan Budi Siswa	1	1
Yayasan Nurul Islam Untung Jawa	1	0
Yayasan Pendidikan Al - Ubadah	0	1

Domus Musculi ARTificialReef, Yayasan Terumbu Rupa	1	1
Yayasan Sayap Ibu Jakarta	1	1
Yayasan Usaha Mulia	1	1
Yayasan Amal Mulia Indonesia	1	1

peluang benar dr uclassify utk mengklasifikasi apakah npish atau bukan adalah 71,43%.

#Contoh panti asuhan disajikan

The number of all potential institutions that we get in the databases are xxxx institutions. The breakdown of them by institution types can be shown in Table x.

#Tabel jumlah institusi disajikan

The potential institutions are clustered by using K-means algorithm to determine whether an institution is NPISH. Its result is shown below:

#Tabel atau gambar hasil K-means

We got xxx institutions which comply the criteria of NPISH.

IV. Conclusion

Based on the results, we have several findings outlined as follows:

-Google Place API is very potential database in order to improve and complement NPISH directory. Moreover, it can be utilized as embryo of NPISH directory.

-To run the clustering and matching process, there are available tools such as uClassify and Excel add ins which can be easily operated.

Nevertheless, we still have some issues to be overcome. First, we should conduct data training for more institutions to increase the accuracy rate of prediction. Second, we should get list of all NPI controlled by government so that we construct the NPISH directory more straightforward and get better results.

V. References

- [1] Badan Pusat Statistik (2014). *Pedoman Penyusunan PDRB Kab/Kota Tahun Dasar 2010 Menurut Pengeluaran*. Jakarta: Badan Pusat Statistik.
- [2] Badan Pusat Statistik (2017). *GDP by Expenditure*, www.bps.go.id/subject/169/produk-domestik-bruto--pengeluaran-.html#subjekViewTab3, accessed on January 30th, 2018.
- [3] Stark, Harold (2017). Since You Asked, Here's How Google Maps Really Works. *Forbes Media*. <https://www.forbes.com/sites/haroldstark/2017/04/26/since-you-asked-heres-how-google-maps-really-works/#7c22eeb4dbed>, accessed on April 2nd, 2018
- [4] United Nations (2009). *System of National Accounts*. New York: United Nations.
- [5] United Nations (2000). *Classification of Expenditure to Purpose: COFOG, COICOP, COPNI, COPP*. New York: United Nations.
- [6] United Nations (2003). *Handbook on Non-Profit Institutions in the System of National Accounts*. New York: United Nations.