

Standardization Problem of Author Affiliations in Citation Indexes

Zehra Taşkın and Umut Al

Hacettepe University, Department of Information Management, 06800, Beytepe ANKARA

Tel: +90 (312) 297 8200

Fax: +90 (312) 299 2014

ztaskin@hacettepe.edu.tr

Abstract: Academic effectiveness of universities is measured with the number of publications and citations. However, accessing all the publications of a university reveals a challenge related to the mistakes and standardization problems in citation indexes. The main aim of this study is to seek a solution for the unstandardized addresses and publication loss of universities with regard to this problem. To achieve this, all Turkey-addressed publications published between 1928 and 2009 were analyzed and evaluated deeply. The results show that the main mistakes are based on character or spelling, indexing and translation errors. Mentioned errors effect international visibility of universities negatively, make bibliometric studies based on affiliations unreliable and reveal incorrect university rankings. To inhibit these negative effects, an algorithm was created with finite state technique by using Nooj Transducer. Frequently used 47 different affiliation variations for Hacettepe University apart from “Hacettepe Univ” and “Univ Hacettepe” were determined by the help of finite state grammar graphs. In conclusion, this study presents some reasons of the inconsistencies for university rankings. It is suggested that, mistakes and standardization issues should be considered by librarians, authors, editors, policy makers and managers to be able to solve these problems.

Keywords: *Standardization Problem, Finite State Technique, Data Accuracy, Data Unification, Address Unification, Research Evaluation, University Rankings, Citation Indexes, Nooj*

Introduction

Citation indexes are used not only for following literature, but also making citation analyses. Citation analysis studies are conducted to measure intellectual effects of researchers and quality of papers (Cole 2000). The content of citation indexes has been growing with the diffusion of the usage of these indexes. In time, some problems of data accuracy have emerged (Galvez & Moya-Anegón 2006a). The data accuracy issues depend on spelling, translation or abbreviation of affiliations (Galvez & Moya-Anegón 2007a). Mistakes originating from affiliations cause problems about showing and evaluating collaborations, limiting scientific fields and effecting performance evaluations negatively (Galvez & Moya-Anegón 2006a).

Scientific studies include organizational and geographical address information of author(s) at the beginning as footnote. In the beginning, address information was given to provide connection between authors and readers. However, the usage of this information has changed gradually with the development of research evaluations and affiliations have become vital for departments, laboratories and research units (De Bruin & Moed 1992). The process of giving affiliations have begun with the author(s)' choice and continued with the formalization of these addresses by the editors' and publishers'. However, when it is left to people's choices, it creates confusion about addresses. In consequence, people working in the same university or department may give different addresses from each other causing hundreds of variations for a university or an organization name in citation indexes. This can lead to serious confusions (Moed 2005: 183-184).

Some organizations' budgets have been determined based on their publication counts. However, since some publications disappear because of address mistakes, organizations or research groups end up losing their budget supports. The situation in Turkey is the same as it is in the world. Organizations have been appraised by using their publication counts. The universities with higher number of publications have been approved as better universities by some authorities. Author(s) are required to publish certain number of articles in citation indexes to take academic degrees (Öğretim 2007). This actually brings the quantitative evaluations to the forefront instead of qualitative ones. On the other hand, Turkish Scientific Research Council (TÜBİTAK) has announced to give support only to articles that have "Turkey" on the address field within the context of incentive program for scientific publications (ULAKBİM 2010). In addition, the rankings of Turkish Universities are declared by The Council of Higher Education every year (The Council of Higher Education 2010). Similarly, URAP (University Ranking by Academic Performance) research laboratory publishes university rankings every year by using various criteria (URAP 2011).

Mentioned implementations have shown the usage of citation indexes in Turkey. Although the numbers do not measure the quality of publications, it is obvious that they have an importance for some communities and policy makers. However, it should be kept in mind that it is unavoidable to make mistakes when using manual indexing systems for citation indexes. Managers should take into account the quantitative analyses based on inaccurate data since access to all publications of each organization has become more of an issue.

The main aim of this study is to develop an algorithm to find mistakes for Turkish Universities in *Web of Science*. First of all, the types of mistakes are identified and their effects are measured. Then, the mistakes are found easily by using an algorithm created by

finite state technique, which has been widely used for recognition of characters, grammar checking, pattern matching, spelling correction and many different areas in the literature. Finite state is defined in the literature as the operation of sets of strings or sequences of a word (Roche & Schabes 1996: 1). Detailed information about finite state technique is explained in the following part. After finding mistakes and standardization errors for affiliations by finite state technique, some suggestions to reduce the problem are given at the end of the study.

The main hypothesis of this study is that “the mistakes in citation indexes can be detected by using finite state technique”. The other hypotheses are as following:

- ◆ There is a standardization problem for Turkish Universities in citation indexes.
- ◆ It is possible to reduce standardization problems by using finite state technique.

Although there are a few publications about data unification for citation indexes in Turkey, this study is the first to identify addresses automatically. Therefore, this study is expected to present some solutions for libraries and decision makers.

Finite State Technique

Finite state algorithms accept strings by following predetermined labels if it can trace a path from the initial state to the finite (Galvez & Moya-Anegón 2007a: 9). These algorithms have networks of defined states and links which are labeled (Roche & Schabes 1995: 236-237). The main operation of finite state depends on reading labeled strings from left to right by considering links between states. If the string matches the predetermined label, the automaton moves on the following state. This process continues until it reaches the final state (Galvez & Moya-Anegón 2007a: 9).

Finite state algorithms are used not only for pattern matching and recognition, speech tagging, recognition of handwriting, optical character recognition and encryption algorithms but also in wide range of scientific areas (Roche & Schabes 1997: 227). It is possible to draw a parallel between finite state algorithm and subway turnstile. To make an analogy, closed turnstile is the initial state for finite state algorithm. If a passenger inserts coin, it moves on second state (gate opens). The turnstile moves on the closed state after passengers pass. By the way the system opens only under the condition of inserting coins (Scholl 2008). The system logic for turnstile has resemblance with finite state algorithms from the point of following and identifying states until the process ends.

Finite state transducers are computer software for implementing finite state algorithms to high amount of texts automatically. Transducers produce strings with regard to existing states to control all stems and forms of a word (Goldsmith 1993; Altıntaş 2001). If all the rules about this word are accepted by transducer, the word is accepted as correct. In the contrary case, the word is rejected or accepted partially (Altıntaş 2001). Finite state algorithm is defined as simple and effective model for natural language processing. Phonological, morphological and syntactical analyses, symbolizations and language modeling can be made easily by using these algorithms (Roche & Schabes 1997). However, usage areas have changed and spread to different fields from linguistics in recent years.

Finite state algorithms are used in many areas in the scientific literature such as engineering, linguistics, medicine and librarianship. The main reason for such commonly use is the customization feature of finite state. In the beginning, although it was propounded that using finite state algorithms to identify natural languages like English was impossible (e.g. Chomsky 1964: 21), it is now commonly used for natural language processing and for revealing morphological structure of languages. Many researchers working on finite state indicate that it can be easily used for natural language processing due to its velocity and density (Johnson 1972; Kaplan & Kay 1994; Roche & Schabes 1995; Oflazer 1996; Mohri 1997).

Previous Studies about Data Accuracy in Citation Indexes

Data accuracy has been popular in recent years for library and information science with the growth of citation indexes. Many studies have tried to find out standardization problems and solve them. According to Moed (2005), the main mistakes have been made because of the complexity of the names of authors and organizations. An author's name can be used in many different formats and there can be many authors with the same name. In addition, translating authors' names from different languages (such as Chinese) to English and using nicknames also create confusion with author names. On the other hand, the problem of organization names depends on flexibility of giving addresses. Namely, two scientists working for the same organization may identify their organizations in different ways causing unauthorized variations in organization names.

Changing names have been determined as one of the main problems for organizations (Hood & Wilson 2003). On the other hand, non-standardization of university names has created problems regarding information retrieval and the solution is well-structured unification (De

Bruin & Moed 1990). Organizations and authors can only be evaluated correctly under the condition of using accurate data (Toutkoushian & Webber 2011: 130). Although the common assumption for university names is “University X”, Van Raan (2005) indicated that this was totally wrong. He also suggested address unification to specify the addresses of all universities.

Finite state algorithms are used in library and information science to make information retrieval more effective (Galvez & Moya-Anegón 2006b; Galvez, Moya-Anegón & Solana 2005; Kettunen 2008) and to standardize author and organization names (Galvez & Moya-Anegón 2006a; Galvez & Moya-Anegón 2007a; Galvez & Moya-Anegón 2007b).

Studies about organization names have been carried out for standardizing organizational name array, which is as follows; university name, institute/faculty/research group, department, city, country. These studies did not aim to determine spelling mistakes for organization names (Galvez & Moya-Anegón 2006a; Galvez & Moya-Anegón 2007a). On the other hand, the study on standardizing author names was designed to find different versions of an author name (Galvez & Moya-Anegón 2007b).

Although there are some papers about standardization in the literature (Cornell 1982; Piternick 1982; Williams & Lannom 1981; Ruiz-Pérez, López-Cózar & Jiménez-Contreras 2002; Falahati Qadimi Fumani, Goltaji & Parto 2012), the issue has not been popular in Turkey, yet. The only unification work for Turkey is a published book on national science indicators (ULAKBİM 2007), which presents all the possible variants of Turkish University names.

The dominant trend for international papers on this topic is finding and solving standardization problems for university names. However, they do not focus on word/spelling mistakes. This study is the first research into the identification of standardization problems in Turkey, presenting some solution proposals for Turkish literature.

Methodology

Many organizations have recently been evaluated in terms of their publication counts in citation indexes. However, quantitative evaluations, like counting publications, are problematic because of the mistakes in some fields of citation indexes. First of all, we gathered 198,687 Turkey-addressed publications that were published between 1928 and 2009 and indexed in *Web of Science* (*SCI*, *SSCI* and *A&HCI*). All types of publications (article, proceeding or letter) were included in our dataset. However, due to some inconsistencies in

the publications' information sections, some publications were excluded. These restrictions are as follows:

- ◆ Some publications in citation indexes such as *Middle Eastern Studies* and *Athenaeum-Studi Periodici Di Letteratura E Storia Dell Antichitabazi* does not have affiliation information.
- ◆ Country information for some publications was incorrect.
- ◆ “Turkey” was not included in the address field for some publications.

Address information for authors found in C1 and RP fields of *Web of Science* were evaluated to identify institutional names in citation indexes. A new column named “institution” was created to write unified addresses for each institution by using Excel. For instance, assume a publication with two authors which has such addresses in C1 field:

- ◆ “Istanbul Univ, Res & Applicat Ctr Biotechnol & Genet Engn, TR-34118 Istanbul, Turkey
- ◆ Gaziosmanpasa Univ, Dept Med Biol, Fac Med, Tokat, Turkey”
- ◆ Sch Med, Gazi Univ, Ankara, Turkey

The addresses of this publication was written in the institution column as “ISTANBUL;GAZIOSMANPASA;GAZI”. Thus, all the publications can be classified under their unified affiliation information; department and faculty names are not unified within the scope of this study. Hierarchical order of the addresses was not considered during unification process. If the organization name was indicated in the middle of the string, it was also unified.

Some publications were excluded from this study because of the unspecified addresses like “dept plast & reconstruct surg, ankara, turkey”. In such a case, the author area (AU) has been evaluated to find the specified university that is in Ankara. If the address information cannot be accessed even by using author names, the records of these publications were not evaluated in this study. Similarly, home and working place addresses like “Fecri Ebcioğlu Sokagi (street), Dilek Apt 6-8,1 Levent, TR-34340 Istanbul, Turkey” were determined and left out of scope. However, this situation did not cause a big problem because only 647 records (0.3% of all records) were affected from that exclusion.

Determination of Standardization Problem and Measure its Magnitude

After unification of 198,687 records, the distribution of Turkey-addressed publications among universities was identified. Then, the most productive first 20 universities that have more than 4,000 publications have been chosen for the determination of standardization problems. The

differences rate between correct addresses and errors have been determined for these 20 universities. The correct address of a university was accepted as “Univ X” and “X Univ” (such as Hacettepe Univ, Univ Hacettepe). Some Turkish Universities have a Turkish and an English name such as Orta Doğu Teknik University; Middle East Technical University. In such a case, all possible correct variants were accepted.

To present the effect of errors on addresses, bibliometric collaboration maps were created by using *CiteSpace* (<http://cluster.cis.drexel.edu/~cchen/citespace/>). Two collaboration maps were drawn to show the effects. First map includes *Web of Science* affiliations (original addresses), and the second one shows unified university names. Then, the differences between these two maps were also evaluated.

Implementing Finite State Transducer

One of the aims of this study is to find all possible variants of a university’s addresses from a huge amount of dataset by using finite state transducer. To achieve this, Hacettepe University, the most productive university in Turkey, was chosen for implementation. Nooj created by Max Silbertzein in 2002 was used as a transducer. Nooj is an open-source linguistic development environment with its large-coverage dictionaries, grammars and corporas (Nooj 2012).

The .txt file that included address data was converted into Nooj text format and named as .not. Then, grammar graphs were created to implement .not file. Detailed information about implementing and creating graphs is explained in the “findings”.

Findings

A total of 198,687 papers were sorted by institution. After the sorting process, the most productive top 20 universities and their publications with the number of incorrect address information are listed in Table 1.

Table 1: The most productive top 20 universities

University	Number of publications	Number of mistakes	Percent of mistakes (%)
Hacettepe University	19,166	340	1.8
İstanbul University	16,390	1,691	10.3
Ankara University	13,275	224	1.7
ODTÜ	11,201	102	0.9
Ege University	9,428	654	6.9
Gazi University	9,281	85	0.9
İstanbul Teknik University	8,613	215	2.5
Dokuz Eylül University	6,069	210	3.5
Atatürk University	5,816	46	0.8
GATA	5,300	639	12.1
Marmara University	5,136	136	2.6
Çukurova University	4,953	60	1.2
Erciyes University	4,795	48	1.0
Ondokuz Mayıs University	4,537	173	3.8
Başkent University	4,418	32	0.7
Selçuk University	4,380	101	2.3
Boğaziçi University	4,102	57	1.4
Fırat University	4,091	86	2.1
Karadeniz Teknik University	4,014	395	9.8
Bilkent University	3,950	11	0.3

Data source: Thomson Reuters Web of Science (<http://isiknowledge.com/wos>)

As it is seen in Table 1, ratios of mistakes differ across universities. Although the most productive institution, Hacettepe University, has lower mistakes than others, second productive İstanbul University lost over 10% of its publications. The main reason of lower loss of Hacettepe University's publications can be explained with the list presented by Hacettepe University Libraries that includes all possible address variants of the university (Hacettepe University Libraries 2012).

Although İstanbul University is announced as the top productive by The Council of Higher Education every year (The Council of Higher Education 2010), it is determined in this study that this university takes place at the second rank with its loss of 10.3%. It is obvious that the publications produced by Istanbul University should be evaluated deeply. In this sense, a search with the keyword "Univ Istanbul" was carried out on April 2, 2012 and 4,383 of the publications did not belong to Istanbul University. Results showed that 2,000 of them were produced by İstanbul Technical University. Result list also included universities that are located in İstanbul such as Koç, Sabancı and Marmara Universities. The main reason of this problem is the addresses that were given as "Koc Univ, Istanbul". Even though searches are conducted with quotation marks, *Web of Science* retrieves these records for "Univ Istanbul"

search. Under these circumstances, both İstanbul University Library and their decision-makers should take this situation into account during ranking and policy making processes.

The most common mistakes were identified for GATA (Gülhane Military Medicine Academy) out of 20 universities. GATA generally takes place at lower ranks in the lists that include publication counts (The Council of Higher Education, 2010; ULAKBİM, 2007: 168). In fact, GATA should be ranked 10th. It is conceivable that all the publications of GATA cannot be determined in the previous studies.

Loss of the publications increased towards the end of the list. The mistake rates of Kahramanmaraş Sütçü İmam (KSU) and Yüzüncü Yıl Universities (YYU), which were not among the top 20 universities, were very high. One fourth of KSU's publications were missing due to the address mistakes. Likewise, YYU lost 15% of its publications.

It is obvious that mistakes make evaluation processes for universities harder. If the reasons of these mistakes can be identified, the solutions will be found easily. In order to find the reasons, a correlation test was carried out for top 40 universities. However, the results of the correlation test showed no meaningful correlation between mistakes and character count ($r=0.064$, $p=0.699$), word count ($r=0.040$, $p=0.810$) or Turkish character count ($r=0.066$, $p=0.692$) on the university names. The only positive correlation was found between total publication count and mistaken publication count ($r=0.585$, $p<0.001$), but this kind of relationship is usually expected and it is natural. Because of these unexplainable errors on addresses, fixing the standardization problem became more problematic.

Error Types

After evaluation and unification processes, different types of mistakes were identified. Main error types are listed below.

Character or Spelling Errors

The main mistakes were specified as errors originating from keyboard while writing university addresses. According to Damerau (1964), over 80% of spelling errors depend on insertion, deletion, substitution and transposition of characters. The same issues were determined for Turkish Universities' addresses such as "Hacetteppe Univ", "Hacattepe Univ", "Maramara Univ", "Egge Univ", "Inonoii Univ" and "Dukuz Eylul Univ". These kinds of errors were made not only by authors, but also by editors or indexers.

Indexing Errors

Web of Science's indexing logic is based on digitization of sources and indexing on the database manually (Thomson Reuters 2009). However, descriptive manuals of *Web of Science* did not explain the way of indexing clearly. An e-mail message from Thomson Reuters Technical Support Team indicated that the indexers depend on the addresses that are written on original texts. In addition to this, it is indicated that an abbreviation list is being used for some words and word groups such as university, faculty, research center etc. It seems that the natural language indexing and digitization of texts cause most of the errors on the address fields of the records in *Web of Science*.

The most interesting error type is the university names written in English instead of Turkish. In *Web of Science*, four records that belonged to Yüzüncü Yıl University were addressed as "Centennial Univ"; six records of Niğde University addressed as "Univ Nigeria"; 12 records of Fırat University addressed as "Univ Florence"; one record of Başkent University addressed as "Univ Basel"; one record of Gazi University addressed as "Graz Univ" and 23 records of Kocaeli, Koç and Afyon Kocatepe Universities were addressed as "Kochi Univ". An example of this kind of mistakes has been evaluated and presented at Figure 1 and 2. It is guessed that indexer had chosen the affiliation from a drop-down menu. However, there is no information about automatic selection of university names. Due to this kind of errors, collaborations for Turkish universities can be interpreted wrongly. For instance, there is no collaborative publication between Turkey and Nigeria, but because of the aforementioned mistake, an incorrect collaboration can be created. These examples prove that, evaluations based on such search results can create confusion on publication counts. The effective solution for this problem is to download the data from citation indexes and clean the data. Such evaluations will reveal more accurate results after the data cleaning process.

Electrical and Optical Properties of an Organic Semiconductor Based on Polyaniline Prepared by Emulsion Polymerization and Fabrication of Ag/Polyaniline/n-Si Schottky Diode

F. Yakuphanoglu,^{*†} E. Basaran,[‡] B. F. Şenkal,[§] and E. Sezer[¶]

Department of Physics, Faculty of Arts and Sciences, Firat University, 23119 Elazig, Turkey, Department of Physics, Gebze Institute of Technology, 41400, Gebze, Turkey, and Department of Chemistry, Faculty of Arts and Sciences, Istanbul Technical University, Maslak, Istanbul, Turkey

J. Phys. Chem. B, 2006, 110 (34), pp 16908-16913

DOI: 10.1021/jp060445v

Publication Date (Web): August 10, 2006

Copyright © 2006 American Chemical Society

Abstract

Figures

Full Text HTML

Citing Articles

Hi-Res PDF [121 KB]

PDF w/ Links [199 KB]

Your current credentials do not allow retrieval of the full text.

Purchase the full-text

Figure 1: Original text of Firat University addressed publication

Electrical and optical properties of an organic semiconductor based on polyaniline prepared by emulsion polymerization and fabrication of Ag/polyaniline/n-Si Schottky diode

Author(s): Yakuphanoglu, F (Yakuphanoglu, F.); Basaran, E (Basaran, E.); Senkal, BF (Senkal, B. F.); Sezer, E (Sezer, E.)

Source: JOURNAL OF PHYSICAL CHEMISTRY B **Volume:** 110 **Issue:** 34 **Pages:** 16908-16913 **DOI:** 10.1021/jp060445v **Published:** AUG 31 2006

Reprint Address: Yakuphanoglu, F (reprint author), Univ Florence, Fac Arts & Sci, Dept Phys, TR-23119 Elazig, Turkey.

Addresses:

[1] Univ Florence, Fac Arts & Sci, Dept Phys, TR-23119 Elazig, Turkey

[2] Gebze Inst Technol, Dept Phys, TR-41400 Gebze, Turkey

[3] Tech Univ Istanbul, Fac Arts & Sci, Dept Chem, Istanbul, Turkey

Figure 2: *Web of Science* record of the publication shown in Figure 1.

Another indexing error type is the mistyping of characters. To illustrate, some addresses have “rn” instead of “m”; “m” instead of “in”; “l” instead of “i”; “i” instead of “l” such as “Parnukkale Univ” (Pamukkale University), “Dokuz Eylui Univ” (Dokuz Eylül University), “F1rat Univ” (Firat University) and “Dumlupmar Univ” (Dumlupınar University). These types of errors may be originating from OCR process of documents. Therefore, digitized materials should be controlled effectively.

Another interesting indexing mistake for Turkish University names is the usage of author surname as affiliation. The publication that was written by “Uslu, G.” and “Uslu, M.” is indexed as “Uslu Univ” in *Web of Science*, even though the original article has correct affiliation (see Figure 3 and 4).

Acne: prevalence, perceptions and effects on psychological health among adolescents in Aydin, Turkey

Author(s): Uslu, G (Uslu, G.)^{1,1}; Sendur, N (Sendur, N.)^{1,1}; Uslu, M (Uslu, M.)^{1,1}; Savk, E (Savk, E.)^{1,1}; Karaman, G (Karaman, G.)^{1,1}; Eskin, M (Eskin, M.)²

Source: JOURNAL OF THE EUROPEAN ACADEMY OF DERMATOLOGY AND VENEREOLOGY **Volume:** 22 **Issue:** 4 **Pages:** 462-469 **DOI:** 10.1111/j.1468-3083.2007.02497.x **Published:** APR 2008

Reprint Address: Uslu, M (reprint author), Uslu Univ, Fac Med, Dept Dermatol, TR-09100 Aydin, Turkey.

Addresses:

[1] Adnan Menderes Univ, Fac Med, Dept Dermatol, Aydin, Turkey

[2] Dept Psychiat, Aydin, Turkey

Figure 3: *Web of Science* record of an incorrect affiliation: “Uslu Univ”

Acne: prevalence, perceptions and effects on psychological health among adolescents in Aydin, Turkey

G Uslu¹, N Şendur¹, M Uslu^{1,*}, E Şavk¹, G Karaman¹, M Eskin²

Article first published online: 20 DEC 2007
DOI: 10.1111/j.1468-3083.2007.02497.x
© 2007 The Authors

J EADV
Journal of the European Academy of Dermatology and Venereology
Volume 22, Issue 4, pages 462–469, April 2008

Additional Information (Show All)

[How to Cite](#) | [Author Information](#) | [Publication History](#)

Author Information

1 Adnan Menderes University Faculty of Medicine, Department of Dermatology,
2 Department of Psychiatry, Aydin, Turkey

*Corresponding author, M. Uslu University Faculty of Medicine, Department of Dermatology, 09100 Aydin, Turkey, tel. + 90 256 4441256/360; fax + 90 2146495; E-mail: meltemozdogan@msn.com

Figure 4: The original article of Figure 3

As seen on Figure 4, although the address of original article is “Adnan Menderes University Faculty of Medicine, Department of Dermatology”, it was indexed in *Web of Science* as “Uslu Univ”. These kinds of mistakes affect the visibility of the publications. In such a case, these publications can only be found by one by one evaluation of all records. This process requires more workforce, time and attention.

Translation Errors Made by Authors

In the international arena, Turkish Universities are addressed by their Turkish names except a few of them, like Middle East Technical University and Istanbul Technical University. *Web of Science* does not translate affiliations into English due to its natural language indexing, yet Turkish authors sometimes prefer to write their affiliations in English. For example, one of

the well-known universities, Boğaziçi University, was indexed in *Web of Science* as “Bosphorus Univ”, but its English name is Boğaziçi University.

Translation for Boğaziçi University does not pose a big problem for this University due to the uniqueness of the name, “Bosphorus”. However, some universities have serious problems because of the translation of their names. For example, some authors used “Aegean University”, “Mediterranean University” and “Trakia University” instead of “Ege University”, “Akdeniz University” and “Trakya University”. This causes confusion since there are other universities bearing that name in the world. In other words, there is a “University of Aegean” in Greece (<http://www3.aegean.gr/>), a “Mediterranean University” in France (<http://www.univmed.fr/>) and a “Trakia University” in Bulgaria (<http://www.unisz.bg/eng/>). Consequently, if someone searches for Ege Universities’ publications and add “Aegean University” to address field, the search results cannot present the correct publication numbers. Obviously, bibliometric studies regarding the number of publications would be inaccurate because of these indexing confusions.

Standardization Problem of University Addresses

Besides the above mentioned errors, standardization of university names is problematic. The problem for university addresses does not only depend on spelling, translating or indexing of the names, but also depends on different usage of university names, such as “X Univ”, “Univ X”, “X Med Sch”, “X Sch Med”, etc. There is no standard array or usage for university names. Galvez & Moya-Anegón (2007a: 8) explained the correct array of university names as “university name, faculty, department, postal code, city, country”. However, most Turkey addressed publications do not have this kind of structure in the affiliations.

The main nonstandard usage of Turkish University names is observed in the abbreviations. For example, Harran and Hacettepe Universities are using HU as abbreviation and this creates confusion about publications origins. To display this, a search was conducted in *Web of Science* by using the terms “HU” and “Turkey” and some results were shown on Figure 5 and 6.

Synthesis, spectroscopy, X-ray structure and redox behaviors of 4-(N-R-salicylideneimine)-2,6-diphenylphenols

Author(s): Kasumov, VT (Kasumov, Veli T.)^{1,1}; Turkmen, H (Tuerkmen, Hasan)^{1,1}; Ucar, I (Ucar, Ibrahim)^{1,2}; Bulut, A (Bulut, Ahmet)^{2,1}; Yayli, N (Yayli, Nurettin)^{1,3}

Source: SPECTROCHIMICA ACTA PART A-MOLECULAR AND BIOMOLECULAR SPECTROSCOPY **Volume:** 70 **Issue:** 1 **Pages:** 60-68 **DOI:**10.1016/j.saa.2007.07.021 **Published:** JUN 2008

Reprint Address: Kasumov, VT (reprint author), **HU**, Fac Arts & Sci, Dept Chem, Sanliurfa, **Turkey**.

Figure 5: HU abbreviation for Harran University in Web of Science

Teachers' proficiency perceptions of about the measurement and evaluation techniques and the problems they confront

Author(s): Gelbal, S (Gelbal, Selahattin)^{1,1}; Kelecioğlu, H (Kelecioğlu, Huelya)^{1,1}

Source: HACETTEPE UNIVERSITESI EGITIM FAKULTESI DERGISI-HACETTEPE UNIVERSITY JOURNAL OF EDUCATION **Issue:** 33 **Pages:** 135-145 **Published:** 2007

Reprint Address: Gelbal, S (reprint author), **HU** Egitim, Fak Egitim Bil Bol Egitimde Olcme Degerlendirm An, Ankara, **Turkey**.

Addresses:

[1] **HU** Egitim, Fak Egitim Bil Bol Egitimde Olcme Degerlendirm An, Ankara, **Turkey**

Figure 6: HU abbreviation for Hacettepe University in Web of Science

Searching with the abbreviations will not retrieve correct results if the organizations do not have unique abbreviations like METU (Middle East Technical University). By searching with the “HU” keyword, one can access the documents written by Harran, Hacettepe, Haliç, Hakkari and Hitit Universities inevitably. In addition to these universities, searching with the “HU” term also brings the addresses like “ICO Badalona, HU Germans, Barcelona, Spain”, “HU Bellvitge, Lhospitalet De Llobregat, Spain”, and “HU Vaudois, Lausanne, Switzerland”. Due to the reasons listed above, the use of abbreviations for universities should be discouraged.

Effects of Mistakes and Non-Standardization

Incorrect and non-standard addresses remarkably affect the accuracy of the search results. There are several problems along with the reduction of visibility of the organizations.

As it is mentioned before, incentive program for scientific publications has been given according to the visibility of country affiliation of the publications in Turkey. If the affiliation is not specified for an article, this article will not have the right to take incentive.

In order to visualize the connections between organizations, and properties of these connections, collaboration networks between organizations are created by bibliometric studies. Such studies need correct, reliable and standard data. Collaboration maps created by

using non-standardized data cannot present the real connections between organizations and they are not meaningful visually, either.

An example for the effects of this kind of data can be seen in Kahramanmaraş Sütçü İmam University's collaboration map. Two maps were created to determine the relation between other universities and Kahramanmaraş Sütçü İmam University. First map was created deliberately by using inaccurate data that comes from an ordinary search in *Web of Science* (see Figure 7).

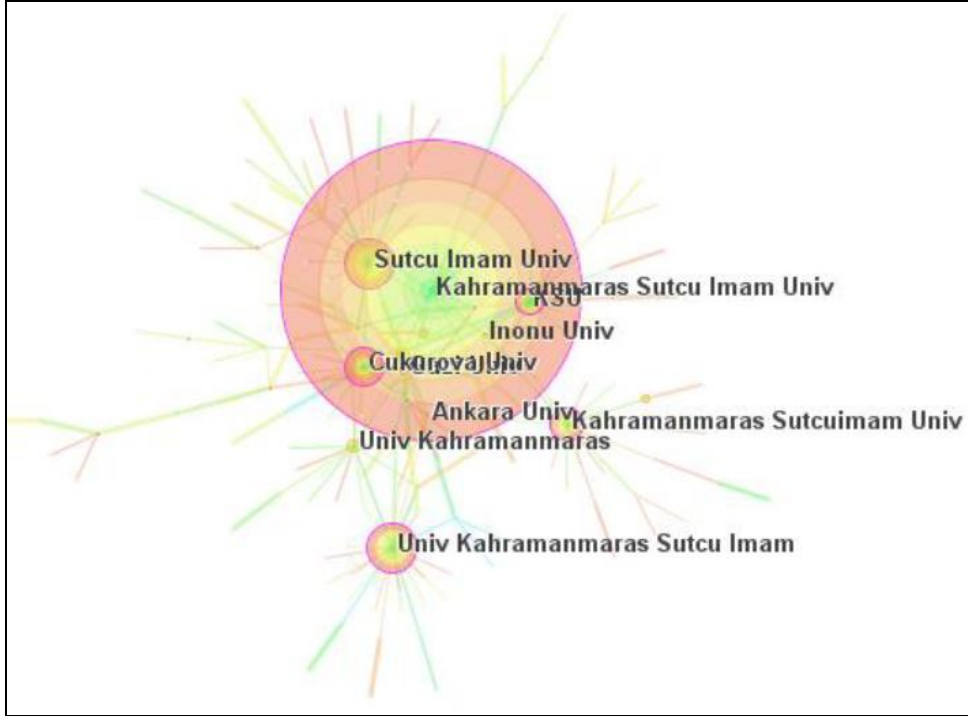


Figure 7: Collaboration map for Kahramanmaraş Sütçü İmam University with inaccurate affiliation information. Collaboration maps are important to show collaborative partners of an institution. However, as it is seen in Figure 7, the connections between institutions are really weak and cannot be determined effectively. This collaboration map should present the collaborative partners of Kahramanmaraş Sütçü İmam University, but many nodes on the map refer to different types of writings of aforementioned university, like “Sutcu Imam Univ”, “Kahramanmaraş Sutcu Imam Univ”, “KSU”, “Kahramanmaraş Sutcuimam Univ”, “Univ Kahramanmaraş Sutcu Imam”. To be able to create this kind of collaboration map, the raw data coming from the *Web of Science* search should be cleaned and the addresses of the universities should be unified. Accordingly, another data set of unified affiliation information was used to create the second map (see Figure 8). It is obvious that the collaborations between institutions have been represented remarkably well by using cleaned data set.

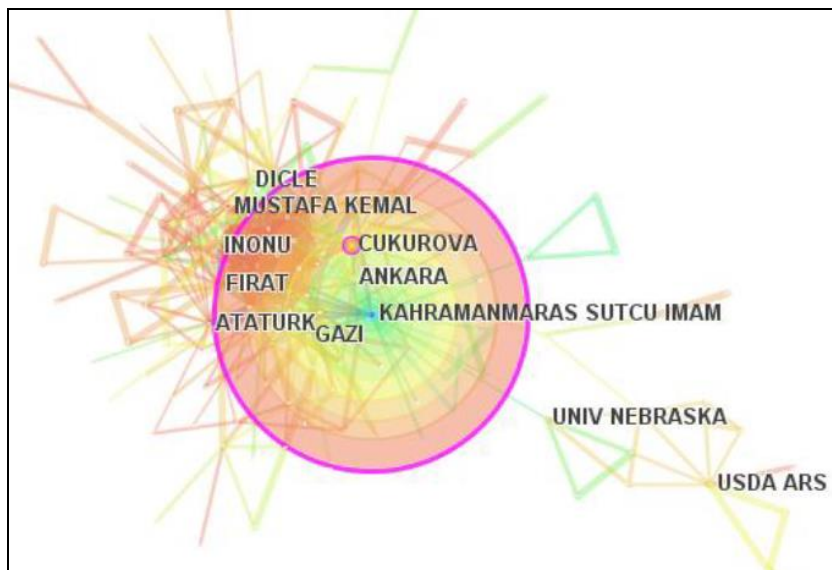


Figure 8: Collaboration map with unified affiliation information

The nodes that cannot be visualized in Figure 7, can be easily seen in the map of unified affiliations (see Figure 8). Figure 8 also shows the major collaborative partners of this university and their connections with each other. The difference between the two figures emphasizes the importance of well-structured unification process. However, working on the unification process manually is time-consuming. If the unification can be achieved by using automatic techniques, the analysis process and the results of bibliometric studies will be easier and far more effective.

Solution Proposals for Standardization Problem

The variety of mistakes and its effects were explained in the previous parts of this study. In this part, the solution proposal for the standardization problem by using the finite state technique and Nooj finite state transducer is introduced.

At first, a .not file was created by using the address fields of all Turkey addressed publications (see Figure 9). The main aim of creating .not file is finding the correct and erroneous

addresses. To achieve this, Hacettepe University was chosen as an example.

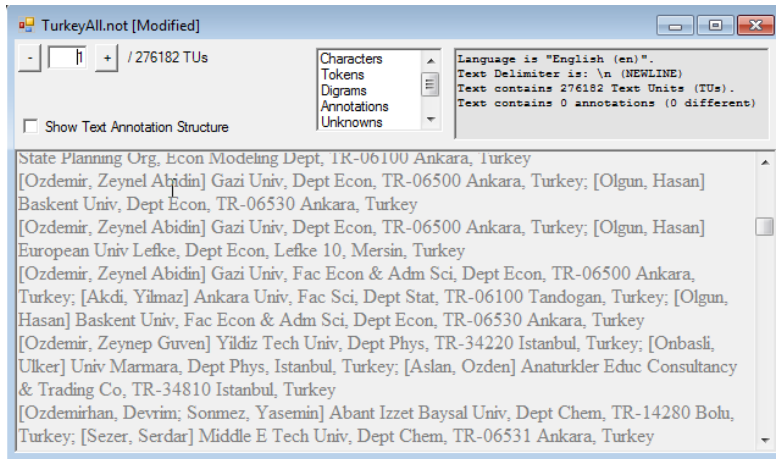


Figure 9: .not file for Turkey-addressed publications

First Stage: Detection of Erroneous Addresses

A finite state graph was drawn to detect address variations for Hacettepe University by using Nooj's File > New > Grammar path. The primary purpose of finite state graph is to identify all possible variants of university's name. Grammar type for the first stage was defined as "productive morphology" and the correct coding for Hacettepe was done. Figure 10 was designed to bring correct variants of "Hacettepe" term.

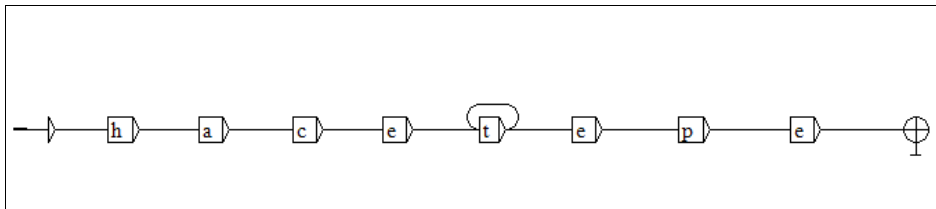


Figure 10: Finite state graph to find correct variants of "Hacettepe" term

In this method, finite state algorithm searches all states from left to right and brings matched records. The reason of circle on "t" is the duplication of "t" character on "Hacettepe" term. This means when the algorithm comes to "t", the character will be repeated. To find out erroneous records, the graph should be enhanced. As mentioned before, main mistakes depend on insertion, deletion, substitution and transposition of characters. Firstly, a graph was developed to find the missing and extra characters and transposition mistakes (Figure 11).

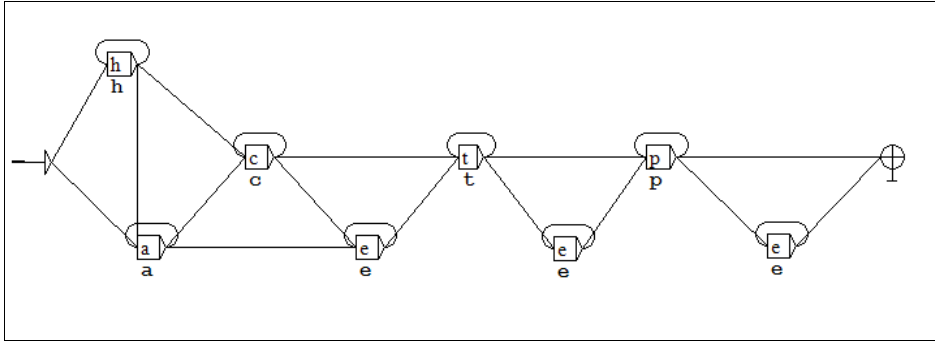


Figure 11: Graph for finding missing and extra characters and transposition mistakes

The circles on the characters work to find extra characters. With this method, even “hhaacceettteeppe” term can be retrieved. Bridges between characters help to find the terms with missing characters. For instance, algorithm can find “hacttepe” term by the help of the bridge between “c” and “t”. Although the beginning state was specified as “h” at first, it is changed into “h” and “a” in the second graph to access the first-letter-missing records.

There are also keyboard mistakes in citation indexes and it should also be added to the graph. Potential keyboard mistakes can be based on character locations. For example, author or indexer can write the name as “Haxettepe” because “c” is close to “x” on Turkish keyboard. Therefore, closest letters on keyword for each state are added to graph. The latest version of first stage’s finite state graph can be seen in Figure 12.

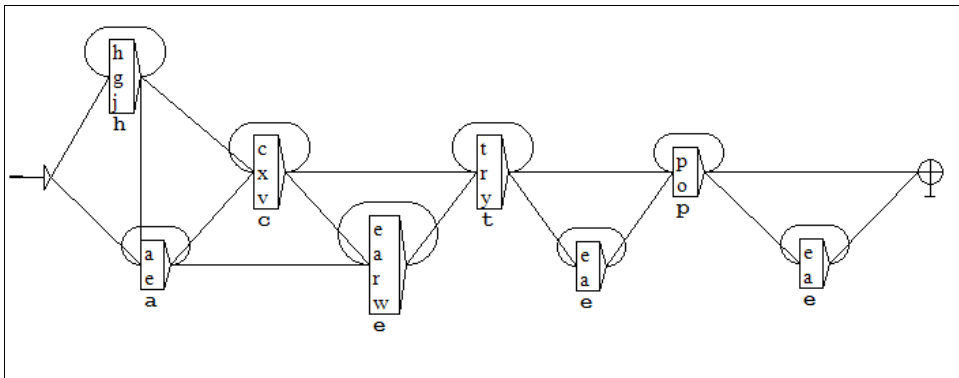


Figure 12: The latest version of finite state graph for first stage

Finite state graph shown on Figure 12 is implied to .not file. Thus, it makes possible to find out erroneous variations of “Hacettepe” term. After morphological analysis, unambiguous words button is used to show identified records. Matching result list is shown on Figure 13.

Text	Before	Seq.	After
[Abaci, Serdar; Yildiz, Attila]	Hacettepe		Univ, Dept Chem, Fac Sci
Albadine, R.; Baydar, Ertoy D.]	Hacettepe		Univ, Ankara, Turkey [Abdullazade, S
Albadine, R.; Baydar, Ertoy D.]	Hacettepe		Univ, Ankara, Turkey [Acar, E
M.; Dogan, B.; Karabudak, R.]	Hacettepe		Univ, Fac Med, Sci Res
Kirazli, Serafettin; Ozcebe, Osman I.]	Hacettepe		Univ, Sch Med, Dept Internal
Hematol, Ankara, Turkey [Acar, Muhittin]	Hacettepe		Univ, Dept Polit & Publ Adm
Athens, GA 30602 USA [Acar, Muhittin]	Hacettepe		Univ, Dept Polit & Publ Adm
[Acar, Ozge C.; Gokmen, Vural]	Hacettepe		Univ, Dept Food Engn, TR
TR-06800 Ankara, Turkey [Acaroglu, Em...	Hacettepe		Univ, Fac Med, Dept Orthoped
TR-06100 Ankara, Turkey; [Bozkurt, Go...	Hacettepe		Univ, Dept Neurosurg, TR-06100 Ankara
Francisco, CA 94143 USA [Acaroglu, Em...	Hacettepe		Univ, Fac Med, Dept Orthoped
Orhan; Saygi, Serap; Ciger, Abdurrahman]	Hacettepe		Univ, Sch Med, Dept Neurol
Ekinci, Saniye; Senocak, Mehmet Emin]	Hacettepe		Univ, Fac Med, Dept Pediat
TR-06100 Ankara, Turkey; [Haliloglu, Mit...	Hacettepe		Univ, Fac Med, Dept Pediat

Figure 13: Concordance for matching results

The concordance shows that there are 27,725 words that matched our dataset. The list of accessed words and their frequencies are presented in Table 2.

Table 2: Accessed words and their frequencies

Term	Freq.	Term	Freq.
Hacettepe	21,024	Hacateppe	2
HACETTEPE	4,658	HACCETEPE	2
HACETEPPE	9	HACCETTEPE	2
Hacattepe	7	HACETTPE	2
HECETTEPE	5	Hacttepe	1
HACETTEPPE	4	Haccattepe	1
Hacetepe	3	Haccettepe	1
Hecettepe	3	HACETEPE	1

As it is seen in Table 2, the total frequency (27,725) is higher than the total publication count of Hacettepe University (19,166). The main reason is the existence of two “Hacettepe” words in the name for some records. In addition, although total mistakes for Hacettepe University were identified as 340, there are only 43 mistakes shown on the Table 2. Rest of the other mistakes depends on standardization problems.

Although all possible variations are tried to be envisaged, the graph still could not retrieve some words (that have undefined errors). However, it is easy to find the unidentified words with the token link on .not file. These words are “Halettepe”, “Hakettepe”, “Hacehepe”, “Hacette” and “HACETIEPE”.

Second Stage: Detection of Unstandardized Addresses

After creating the grammar for spelling mistakes, another grammar has been developed to identify the variety of addresses apart from “Hacettepe Univ” and “Univ Hacettepe”.

The second stage is about syntactic rules for the Hacettepe University and consequently “syntax” module was chosen for the second stage.

The main aim of the second stage is to create a graph that reveals the terms used with the word “Hacettepe”, since “univ” abbreviation is not enough to access all the documents which refer to this university. Finite state graph that aims to detect unstandardized addresses is presented on Figure 14.

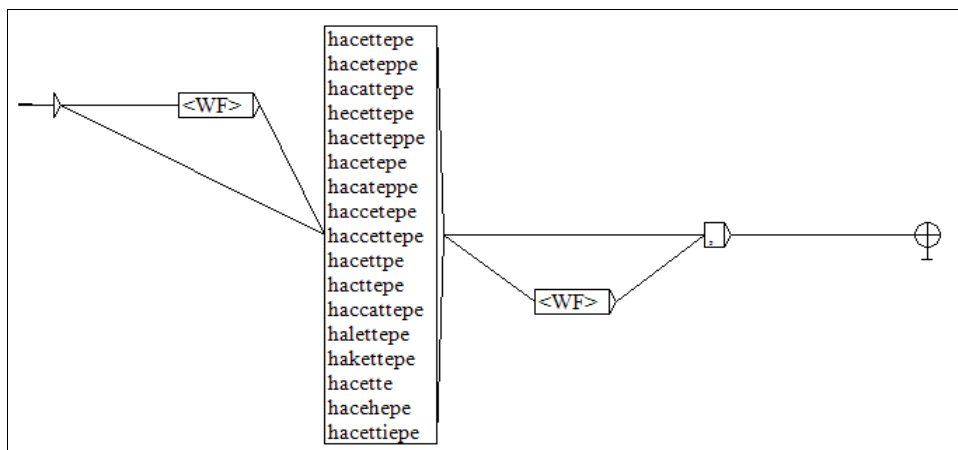


Figure 14: Finite state graph that detects unstandardized addresses for Hacettepe University

All addresses in the address field of *Web of Science* are divided into categories (faculty, department, city etc.) followed by a comma. Therefore, the graph is ended with a comma to determine only the university name. Any words that are placed before and after the term Hacettepe can be retrieved by this algorithm. After implementing the graph into .not file, Nooj produced a concordance which is shown on Figure 15.

Before	Seq.	After
CHILDRENS HOSP,ANKARA 06100,TURKEY	HACETTEPE UNIV,	FAC MED,DEPT PEDIA
T,HEMATOL UNIT,ANKARA 06100,TURKEY;	HACETTEPE CHILDRENS HOSP,	ANKARA 06100,TURK
CHILDRENS HOSP,ANKARA 06100,TURKEY	Hacetepe Univ,	Dept Food Engr, Fac En
Fac Engr, TR-06532 Ankara, Turkey	HACETTEPPE CHILDRENS HOSP,	ANKARA,TURKEY HA
EPPE CHILDRENS HOSP,ANKARA,TURKEY	HACETTEPPE UNIV,	ANGORA,TURKEY HA
MED,DEPT PEDIAT,ANGORA,TURKEY	HACETTE UNIV,	DEPT ENGLISH LANG
Dept Cardiol & Hematol, Ankara, Turkey	HACETTEPE CHILDREN HOSP,	ANKARA,TURKEY HA
HOSP MED CTR,ANGORA,TURKEY;	HACETTPE UNIV,	SCH MED,DEPT PEDIA
FDN,PEDIAT UNIT,ANKARA,TURKEY;	HECETTEPE CHILDRENS HOSP,	DEPT PEDIAT SURG,A
Med, Dept Microbiol, Ankara, Turkey	HACETTEPE FAC MED,	DEPT PLAST & RECON
Med, Dept Ophthalmol, Ankara, Turkey	Hacettepe Hastanesi,	Dept Ophthalmol, TR-06
Dept Nephrol & Hematol, Ankara, Turkey	HACETTEPE HOSP,	DEPT PEDIAT CHEST I
Gastroenterol Unitesi, TR-06100 Ankara, Turkey	Hacettepe Inst Oncol,	Ankara, Turkey; Brigham
Pediat Rheumatol & Nephrol, Ankara, Turkey	Hacettepe KUniv Hastaneleri,	Endokrinol Bolumu, Hosp
Dept Endocrinol & Metab, Ankara, Turkey	HACETTEPE MED CTR,	ANKARA,TURKEY HA
Fac, Dept Hematol, Ankara, Turkey	HACETTEPE MED FAC,	ANESTHESIOLOG MAIN
CH,DEPT PAEDIAT,ANKARA 06100,TURKEY	Hacettepe Oncol Inst,	Ankara, Turkey Hacettep
Fak, Psikiyatri AD, Ankara, Turkey	Hacettepe U,	Ogrenci Saglik Merkezi, J
UNIV,EYE CLIN,ANGORA,TURKEY	HACETTEPE UNIV BEYTEPE,	DEPT CHEM,ANKARA
IST ONCOL,ANKARA,TURKEY HACETTEPE	UNIV HACETTEPE MED SCH,	DEPT PEDIAT,BLOOD
HOSP,BURN UNIT,ANKARA,TURKEY	HACETTEPE UNIV MED,	DEPT GEN SURG,ANK

Figure 15: Final concordance for Hacettepe University

It is specified that there are 47 different usages apart from “Univ Hacettepe” and “Hacettepe Univ” that refers to Hacettepe University. All detected addresses are shown on Table 3.

Table 3: Detected addresses for Hacettepe University

Term	Freq.	Term	Freq.
Hacettepe Univ / Univ Hacettepe	18,826	Haccetepe Fac Med	1
Hacettepe Childrens Hosp	84	Haccettepe Univ Hosp	1
Hacettepe Med Sch	60	Hacettepe Med Acad	1
Hacettepe U	33	Hacettepe Kuniv Hastaneleri	1
Hacettepe Sch Med	28	Hacehepe Univ	1
Hacettepe Med Fac	23	Hacettepe Childrens Hosp	1
Hacettepe Fac Med	12	Hacette Univ	1
Hacettepe Hastaneleri	9	Hacettepe Technopolis	1
Hacettepe Hosp	9	Sociales Hacettepe	1
Hacettepe Hastanesi	7	Hacettepe Adult Hosp	1
Hacettepe Oncol Inst	5	Hacettepe Child Hosp	1
Hacettepe Univ	5	Hacettepe Cocuk Hastabanesi	1
Hecettepe Univ	5	Hacettepe Univ	1
Hacettepe Med Ctr	4	Hacettepe Childrens Hosp	1
Hacettepe Tip Fak	3	Hacettepe Univ	1
Hacettepe Univ	3	Univ Hacettepe	1
Hacettepe Children Hosp	3	Hacettepe Cocuk Hastahanesi	1
Hecettepe Univ	3	Hacettepe Cocuk Hastanesi	1
Hacettepe Univ	2	Hacettepe Cocuk Hastanesi	1
Hacettepe Univ	2	Hacettepe Eriskin Hastanesi	1
Hacettepe Univ	2	Hecettepe Childrens Hosp	1
Hosp Hacettepe	2	Klinikum Hacettepe	1
Hacettepe Univ Hosp	1	Hacettepe Inst Oncol	1
Haccattepe Univ	1	Unit Hacettepe	1

Finite state graphs could not retrieve 10 different addresses (such as “Laacettepe Univ”, “Ihsan Dogramaci Childrens Hosp”, “HUTF Plast Cerrahi ABD”) that are used only 10 times. In previous studies (ULAKBIM, 2007: 354-355), 69 different addresses for Hacettepe University were identified in *Web of Science*. Although, unretrieved addresses (“Ihsan Dogramaci Childrens Hosp” and “HU Biol Dept”) and the 21 of the retrieved addresses that contained “Hacettepe Univ” or “Univ Hacettepe” (like “Hacettepe Univ Hastaneleri”, “Hacettepe Univ Hastanesi”, “Hacettepe Univ Med”) were covered. 11 records which were accessed by finite state technique did not take place in the list of ULAKBIM. As a result, 49 out of 59 different address variations could be accessed by the graph and they are the ones which were frequently used in the address field of *Web of Science*.

It is concluded that identifying and accessing address variations for universities is possible by using the methodology of this study. However, it is hard to apply this technique for the

universities like Ege and Gazi because of the characteristics of their names. It is possible to say that this technique can be applied to universities that have a distinctive name.

Conclusion

Some quantitative analyses based on publication counts of universities and organizations have been commonly used and taken into consideration by some authorities. Therefore, the general opinion about publications has been transformed to “more publication indicates better organizations”. Moreover, the existence of publications in citation indexes is becoming more and more prominent, which indicates that citation indexes’ main aim of usage has been changing dramatically. Attaching a particular value to publication counts makes it more important to determine all publications for each organization. However, calculating publication counts has been problematic because of the manual indexing in citation indexes. Some indexing mistakes have made all evaluations depending on publication counts unreliable.

International visibility is vital for some organizations to catch collaboration opportunities. It is also important to create correct collaboration maps to represent the networks between organizations. The lack of standardization has not only been affecting quantitative analyses, but also reducing institutional visibility of universities and organizations. Quantitative analyses have been also popular for Turkey and they have been affecting public opinion about universities recently. However, many of these evaluations present different results from each other because of the inaccurate data. It is quite obvious that making standardization is quite important for reflecting correct results with bibliometric studies.

The mistaken affiliations in citation indexes for Hacettepe University have been specified with this study. Also, the finite state technique is proposed to standardize affiliations by designing some finite state algorithms. The main hypothesis which was determined as “the mistakes in citation indexes can be detected by using finite state technique” is proved at the end of the study. This technique can work for many universities which have distinctive names such as Hacettepe, Uludağ, and Atatürk Universities. However, as mentioned in findings part, it can be foreseen that this technique is not applicable for short-named organizations (like Ege).

Unidentified address variations of Hacettepe University can be retrieved with the designed finite state algorithm which had been missed by the previous studies. Consequently, the effective results can be obtained by using finite state algorithms with least effort for annual

publication count reports of Turkey. Furthermore, a general algorithm can be developed as a future study to extract all possible address variants for all Turkish Universities.

This study also shows the accuracy and reliability problems of citation indexes and quantitative rankings. The policies developed by using publication counts can be unreliable in parallel with the questionable data of citation indexes. Moreover, evaluating universities' performances with quantitative methods should be investigated. Future studies can comprise alternative evaluation methods instead of counting publications.

Suggestions

The determination, reduction and evaluation of mistaken records have been more important for librarians, authors, editors, policy makers and managers. In this part, there will be some suggestions for each group.

- University libraries should be conscious of mistaken data when reporting statistics to authorities. Searching citation indexes by using some basic search terms and use all the gathered records for reporting is not a perfect way to represent the real performance. Downloading data from citation indexes and cleaning it is a better way to access correct records than searching. Libraries can also suggest some standard alternatives about possible usage of the institution's affiliations on their web sites. Hacettepe University has such guidance on its web site and by this way its loss on publication count seems lower when compared to most of the universities in Turkey.
- Authors should be careful when they write affiliations on their studies. If any mistaken affiliation is detected for a study which has the correct address in the original text, that means it is possibly an indexing mistake and, it can be corrected by Thomson Reuters. In such a case, authors can fill the correcting form and follow the process in the web site (<http://ip-science.thomsonreuters.com/techsupport/datachange/>).
- The job also falls to the editors on the process of formalization of affiliations. They should review the articles properly and correct the affiliation mistakes. It is presumable that there will be lower mistakes for publications that are evaluated deeply during the editorial process.
- The providers of citation indexes have the main duties about address standardization as they are the critical actors in the field. The providers may lost their confidence and prestige in the community. Therefore, to leave the manual indexing should be their primary task, since it is hard to identify the human-induced mistakes than the mistakes

originated from computers. One of the well-known databases, *Scopus*, has challenged this issue with its identifier mechanism entitled “affiliation identifier” (SciVerse Scopus 2012). As for *Web of Science*, there are some works to unify author names (ResearcherID) and affiliations (Organization-Enhanced list). Thomson Reuters launched “Organization-Enhanced list feature of *Web of Science* in May 2012, which allows users to search preferred organization names and/or their name variants to add their search queries (Thomson Reuters, 2012). Although these efforts offer some quick and practical formulas to solve the problem, it should be taken into account that standardization problem can be minimized only with unique identifiers; all other efforts to solve them generate temporary solutions.

- University rankings are one of the hot topics on the agenda of some organizations. Students decide their schools by checking its rank and Turkish Research Council gives incentives to researchers according to the organizational affiliations of their publications. Therefore, managers and policy makers should consider about data accuracy issues in citation indexes. Not the values of scientific works but their numbers are becoming more and more important for Turkey. It is alarming that if this situation continues, there will be a group of useless publications. The most important thing is to find some ways to determine quality of works.

Acknowledgments

This article is based on Taşkın’s (2012) MA thesis and was supported in part by a research grant of the Turkish Scientific and Technological Research Center (110K044). We thank Dr. İrem Soydal and Dr. Mustafa Şahiner for their meticulous reading of a draft version of this paper and for their invaluable suggestions.

References

- Altıntaş, K. (2001). *Turkish to Crimean Tatar machine translation system*. Unpublished Master’s Thesis, Bilkent University, Ankara.
- Chomsky, N. (1964). *Syntactic structures*. The Hague: Mouton de Gruyter.
- Cole, J. R. (2000). A short history of the use of citations as a measure of the impact of scientific and scholarly work. In *The Web of Knowledge: a festschrift in honor of Eugene Garfield* (pp. 281-298). New Jersey: Information Today.
- Cornell, L.L. (1982). Duplication of Japanese names: a problem in citations and bibliographies. *Journal of the American Society for Information Science*, 33(2), 102-104.
- Damerau, F.J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3), 171-176.

- De Bruin, R.E. & Moed, H.F. (1990). The unification of addresses in scientific publications. *Informetrics*, (89-90), 65-78.
- Falahati Qadimi Fumani, M.R., Goltaji, M. & Parto, P. (2012). Inconsistent transliteration of Iranian university names: a hazard to Iran's ranking in ISI Web of Science. *Scientometrics*, doi: 10.1007/s11192-012-0818-2.
- Galvez, C., Moya-Anegón, F. & Solana, V.H. (2005). Term conflation methods in information retrieval: non-linguistic and linguistic approach. *Journal of Documentation*, 61(4), 520-547.
- Galvez, C. & Moya-Anegón, F. (2006a). The unification of institutional addresses applying parametrized finite-state graphs (P-FSG). *Scientometrics*, 69(2), 323-345.
- Galvez, C. & Moya-Anegón, F. (2006b). An evaluation of conflation accuracy using finite-state transducers. *Journal of Documentation*, 62(3), 328-349.
- Galvez, C. & Moya-Anegón, F. (2007a). Standardizing formats of corporate source data. *Scientometrics*, 70(1), 3-26.
- Galvez, C. & Moya-Anegón, F. (2007b). Approximate personal name-matching through finite-state graphs. *Journal of the American Society for Information Science and Technology*, 58(13), 1-17.
- Goldsmith, J.A. (1993). *The last phonological rule: reflections on constraints and derivations*. Chicago: University of Chicago Press.
- Hacettepe University Libraries. (2012). İndekslerde HÜ-TR: bilimsel yayınlarda adres bilgisi [HU-TR in citation indexes: address information for scientific publications].
<http://library.hacettepe.edu.tr/index.php?kid=105&sid=111>. Accessed 28 December 2012.
- Hood, W.W. & Wilson, C.S. (2003). Informetric studies using databases: opportunities and challenges. *Scientometrics*, 58(3), 587-608.
- Johnson, C.D. (1972). *Formal aspects of phonological description (monographs on linguistic analysis)*. The Hague: Mouton De Gruyter.
- Kaplan, R.M. & Kay, M. (1994). Regular models of phonological rule systems. *Computational Linguistics*, 20, 301-378.
- Kettunen, K. (2008). Reductive and generative approaches to management of morphological variation of keywords in monolingual information retrieval: an overview. *Journal of Documentation*, 65(2), 267-290.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.
- Mohri, M. (1997). Finite-state transducers in language and speech processing. *Computational Linguistics*, 23(2), 269-311.
- Nooj. (2012). Nooj Introduction. <http://www.nooj4nlp.net/pages/introduction.html>. Accessed 28 December 2012.
- Oflazer, K. (1996). Error-tolerant finite-state recognition with applications to morphological analysis and spelling correction. *Computational Linguistics*, 22(1), 73-89.
- Öğretim Üyeliğine Yükseltilme ve Atanma Yönetmeliğinde Değişiklik Yapılmasına Dair Yönetmelik [Regulation for changings on promotions and assignments for lecturers]. *T.C. Resmi Gazete*.
<http://www.resmigazete.gov.tr/main.aspx?home=http://www.resmigazete.gov.tr/eskiler/2007/05/20070526.htm&main=http://www.resmigazete.gov.tr/eskiler/2007/05/20070526.htm>. Accessed 28 December 2012.

- Piternick, A.B. (1982). Standardization of journal titles in databases. *Journal of the American Society for Information Science*, 33(2), 105.
- Roche, E. & Schabes, Y. (1995). Deterministic part-of-speech tagging with finite-state transducers. *Computational Linguistics*, 21(2), 227-253.
- Roche, E. & Schabes, Y. (1996). Introduction to finite-state devices in natural language processing. Technical Report, Mitsubishi Electric Research Laboratories, TR96-13. <http://www.merl.com/papers/docs/TR96-13.pdf>. Accessed 27 February 2012
- Roche, E. & Schabes, Y. (1997). *Finite-state language processing (language, speech and communication)*. Cambridge: The MIT Press.
- Ruiz-Pérez, R., López-Cózar, E.D. & Jiménez-Contreras, E. (2002). Spanish personal name variations in national and international biomedical databases: implications for information retrieval and bibliometric studies. *Journal of the Medical Library Association*, 90(4), 411-430.
- Scholl, M.H. (2008). (Some) Formal foundations of modelling dynamics. <http://www.inf.uni-konstanz.de/fileadmin/dbis/im08/4-Dynamik-B.pdf>. Accessed 28 December 2012.
- SciVerse Scopus. (2012). Affiliation identifier. http://help.scopus.com/flare/schelp_Left.htm#CSHID=h_affilrchtips.htm|StartTopic=Content%2Fh_affilrchtips.htm|SkinName=svs_SC. Accessed 28 December 2012.
- Taşkın, Z. (2012). Atıf dizinlerinde üniversite adreslerinin standardizasyon sorunu (Standardization problem of university addresses on citation indexes). Unpublished MA Thesis, Hacettepe University.
- The Council of Higher Education. (2010). 2010 yılı yayın istatistikleri [Publication statistics of 2010]. http://www.yok.gov.tr/index.php?option=com_docman&task=doc_download&gid=728. Accessed 28 December 2012.
- Thomson Reuters. (2009). Web of Science 8.0. <http://ip-science.thomsonreuters.com/m/pdfs/mgr/ws-wos-8-0-0807.pdf>. Accessed 28 December 2012.
- Thomson Reuters. (2012). Searching the Organizations - Enhanced List. http://images.webofknowledge.com/WOKRS57B4/help/WOS/hp_organizations_enhanced_index.html. Accessed 20 February 2013.
- Toutkoushian, R.K. & Webber, K. (2011). Measuring the research performance of post-secondary institutions. In *University Rankings: theoretical basis, methodology and impacts on global higher education* (pp. 123-144). New York: Springer
- ULAKBİM. (2007). *Türkiye bilimsel yayın göstergeleri [National Scientific Indicators for Turkey]*. İ.H. Demirel, C. Saraç & E.A. Gürses (Eds). Ankara: ULAKBİM.
- ULAKBİM. (2010). TÜBİTAK Uluslararası bilimsel yayınları teşvik programı uygulama esasları [Implementation fundamentals for TÜBİTAK Incentive Program for Scientific Publications]. ULAKBİM. http://www.ulakbim.gov.tr/cabim/ubyt/esaslar_v2.pdf. Accessed 28 December 2012.
- URAP (University Ranking by Academic Performance). (2011). Genel Bilgi [General information]. http://tr.urapcenter.org/2011_2/2011_7.php. Accessed 28 December 2012.
- Williams, M.E. & Lannom, L. (1981). Lack of standardization of the journal title element in databases. *Journal of the American Society for Information Science*, 32(3), 229-233.

Van Raan, A.F.J. (2005). Fatal attraction: conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133-143.