
AI translation · View original & related papers at
chinaxiv.org/items/chinaxiv-202310.00683

Research on the Linked Dataset of Patents and Patent References Based on the ISLI Standard (Postprint)

Authors: Zhu Jiang, Zhang Di, Liu Chunjiang, Zhou Qi

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose/Significance] Utilizing the ISLI standard to establish a linked dataset between patents and patent references promotes bibliometric research and studies on science-technology linkages. [Method/Process] Based on analyzing the linkage scheme between patents and patent references, along with unique identifiers of target literature and their acquisition methods, this paper explores the methodology for constructing such a linked dataset using the ISLI standard. [Results/Conclusion] The linked dataset can enrich and refine the existing citation system, bridge the two major systems of patent and non-patent literature, and enable more accurate measurement of citation counts and impact for both patent and non-patent literature.

Full Text

Study on Linkage Data-Sets of Patents and Patent References Based on ISLI Standard

Zhu Jiang¹², Zhang Di¹², Liu Chunjiang¹², Zhou Qi³

¹Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610041

²Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

³EBSCO International, Inc., Beijing 100026

Abstract

[Purpose/Significance] This paper aims to establish linkage data-sets of patents and patent references based on the ISLI standard, thereby promoting

bibliometric research and studies on the linkages between science and technology. **[Method/Process]** Based on an analysis of linkage schemes between patents and patent references, as well as an examination of unique identifiers for target documents and their acquisition methods, this study explores the methodology for constructing patent-reference linkage data-sets using the ISLI standard. **[Result/Conclusion]** The resulting linkage data-sets can enrich and improve existing citation systems, bridge the two distinct systems of patent and non-patent literature, and enable more accurate measurement of citation counts and influence for both patent and non-patent documents.

Keywords: international standard link identifier; patent; non-patent references; linkage data-set; ISLI

1. Introduction

Patents represent an integrated information resource encompassing technology, economics, and legal dimensions [1], constituting a crucial form of intellectual property. Prominent commercial patent databases include PatSnap, incoPat, Derwent Innovations Index (DII), Derwent Intelligence (DI), Innography, and Orbit, while major non-commercial databases include the European Patent Office's esp@cenet database. Most of these patent databases contain over 120 million patent records. Non-patent literature, by contrast, exists in vastly greater quantities, with discovery systems such as Summon and EDS providing access to billions of metadata records across various document types.

Patent and non-patent literatures respectively represent the outputs of technological development and scientific research, exhibiting close interconnections. These relationships manifest not only through common institutions and authors, but also through thematic continuity, technical lineage, and mutual citation patterns in reference lists.

Patent references comprise two categories: patent references (citations to other patents) and non-patent references (NPR) such as journal articles, books, conference papers, and other non-patent documents [2]. Many scholars treat non-patent and patent literature as proxies for science and technology, respectively, exploring their interconnections through analyses of institutions, authors, subjects, and citations to understand the linkages [3-4], evolution [5], and trends [6] between science and technology. Accurate annotation and full-text linking of both patent and non-patent references in patent databases would significantly facilitate patent examination, enhance understanding of novelty and inventiveness, and advance scholarly research on science-technology linkages through full-text analysis.

2. Current State of Patent-Reference Linkage

Given the close relationship between patent and non-patent literature, researchers strongly desire integrated access. Notable integration systems include Web of Knowledge and InnovationQ Plus. The Web of Knowledge platform's DII database contains approximately 120 million patent records, while Web of Science covers only top-tier non-patent literature, with SCIE indexing over 9,000 scientific journals and the Conference Proceedings Citation Index covering approximately 3,000 conferences annually—representing a relatively limited scale. Although the Web of Knowledge platform enables unified searching across patents and non-patent literature, the search results lack explicit linkages between these document types.

InnovationQ Plus claims to be a “unified search system for patents and non-patent literature,” but primarily serves fields such as electronics and communications rather than covering all scientific and technological disciplines. Most other patent databases include patent references, but while patent-to-patent citations can be conveniently linked via publication or application numbers, non-patent references rarely provide direct links to bibliographic information, let alone full text. This limitation creates significant inconvenience for users seeking to access non-patent references. Therefore, research on patent-non-patent reference linkage is essential to help users quickly and accurately locate and link to the bibliographic information and full text of cited non-patent documents.

3. Types of Patent References

As described above, patent references include both patent and non-patent categories. Using European Patent Office docdb data as an example, non-patent references encompass journal articles, books, chemical abstracts, databases, biological abstracts, serials/journals/periodicals, and World Wide Web sources. Table 1 illustrates the codes, types, and examples of non-patent references in EPO docdb data.

The most common non-patent reference type is code A (abstract citation of no specific kind), which includes journal articles, dissertations, and technical reports. Table 2 presents basic information and sample non-patent references for U.S. Patent US9622820B2.

4. ISLI-Based Linkage Data-Set Design

Establishing a linkage data-set between patents and their references requires connecting two types of documents. Multiple approaches exist, one of which employs the ISO 17316:2015 Information and Documentation—International Standard Link Identifier (ISLI) [7], a standard developed under Chinese leadership and officially published in 2015.

ISLI represents a novel identifier concept that does not identify a single entity but rather the association between two entities [8]. Its function is to establish

links between a “source” and a “target” entity, each bearing specific identifiers, with the entities potentially having either identical or different granularities. This characteristic provides solutions for linking different types of entities [9].

4.1 ISLI Encoding Scheme Using ISLI association encoding, a patent containing references can serve as the “source,” with each cited reference designated as a “target.” Assigning an ISLI code to each source-target pair creates an ISLI association record. The “source” can be identified by publication number, while “targets” may include journal articles, conference papers, dissertations, monographs, patents, and other document types, identifiable by DOI, handle, ISBN-A, URN, or publication number (see Figure 1 [Figure 1: see original paper]).

The ISLI code comprises decimal digits divided into three parts: service field, association field, and check field [10]. For patent-reference linkages, the ISLI code can be structured as follows (see Figure 2 [Figure 2: see original paper]):

The first part is the service field, typically 6 digits, assigned by the ISLI Registration Authority (ISLI RA). The second part is the association field with variable length, subdivided into prefix and suffix codes. Considering current global patent volumes and future growth trends, the prefix code can be set to 10 digits, enabling linkage encoding for 10 billion patents without expansion. The suffix code can be 5 digits, where code 00000 points to the patent’s detailed record, and codes 00001-99999 allow up to 99,999 associations per patent for references or other attributes. The third part is the check field (checksum), automatically calculated by the ISLI system according to specified rules.

5. Acquisition of Reference Identifiers

5.1 Patent Reference Identifiers Within a patent database system, patent references can be easily located and linked via publication number, patent number, or application number. Certain conversion rules even enable direct linking to detailed records and full-text specifications on patent office websites such as the European Patent Office. This patent-to-patent association is relatively straightforward, as both documents can be directly linked using their publication numbers.

5.2 Non-Patent Reference Identifiers Linking non-patent references is considerably more complex due to their diverse types. Analysis reveals that journal articles constitute the majority of non-patent references and are published according to standardized norms, with nearly all available in digital formats online. Open-access articles provide direct full-text access, while commercially published articles allow authorized users to view full text within licensed IP ranges or via login credentials; unsubscribed users can typically access abstracts free of charge.

Using the journal article “Ye X, Liu H, Chen L, et al. Reverse innovative design—an integrated product design methodology. Computer-aided

design, 2008, 40(7): 812-827” from Table 2 as an example, one can query <http://www.crossref.org/guestquery/> using the first author’s surname and article title to obtain the persistent link (<http://dx.doi.org/10.1016/j.cad.2007.07.006>) and DOI (see Figure 3 [Figure 3: see original paper]). Clicking the persistent link redirects to the publisher’s website for the article (see Figure 4 [Figure 4: see original paper]), where users can view detailed information and abstracts, with authorized users able to download the full text. The DOI serves as the article’s identifier.

If permanent links and DOIs cannot be obtained through this method, resource discovery systems offer an alternative. Using the EDS discovery system as an example, its RESTful API-based data retrieval interface returns XML/JSON results. During data processing and system operation, the EDSAPI interface can query the extensive EDS metadata repository using composite conditions such as “article title + journal name + date” to retrieve detailed article information.

For the article “Land use change monitoring in nature reserves based on GF-1/GF-2,” parsing the returned JSON data from EDS (see Figure 5 [Figure 5: see original paper]) yields the article’s URL and other metadata, with the accession number (AN) serving as the identifier. Dissertations, monographs, and other non-patent reference types can similarly obtain handle, ISBN-A, URN, or EDS AN identifiers. Once identified, non-patent references can be associated with patents through ISLI.

6. Implementation of Linkage Data-Set Construction

6.1 Patent Big Data Platform Development The Chengdu Library and Information Center of the Chinese Academy of Sciences has built a patent data service platform (<http://tmlcloud.casip.ac.cn/pbsp/>) using EPO’s XML-format patent data (see Figure 6 [Figure 6: see original paper]). The platform indexes 120 million patent bibliographic and legal status records from over 100 countries, regions, and organizations. Powered by a big data analytics engine, it provides intelligent search, professional analysis, knowledge discovery, and data download functions. Intelligent search combines diverse strategies with semantic retrieval technology and multi-faceted browsing modes. Professional analysis employs multi-dimensional indicators, while knowledge discovery leverages big data computing for technology mining and forecasting. Data download supports customized user requirements.

6.2 Resource Discovery System Implementation The Chengdu Library and Information Center has also deployed the EDS discovery system (see Figure 7 [Figure 7: see original paper]), enabling one-stop discovery across over 100 commercial and open-access databases subscribed to or linked by the center. While these two systems respectively integrate patent and non-patent literature, they previously lacked inter-system linkages.

6.3 Patent Reference Linkage Implementation On the patent data platform, patent-to-patent linkages have been implemented via publication numbers, allowing users to click patent references and navigate directly to detailed patent pages. For non-patent references, specialized programs retrieve permanent links and DOIs for journal articles, writing these into the patent database and displaying them in reference lists. Users clicking these links can access article abstracts on publisher platforms, with authorized users able to download full text. For articles without permanent links, the EDSAPI interface retrieves EDS AN numbers, full-text URLs, and other metadata, which are similarly stored and displayed for user access.

6.4 Linkage Data-Set Generation Generating patent-reference linkage data-sets using ISLI requires assigning each patent one basic ISLI code (e.g., ISLI 012345-012345678900000-8) with 99,999 associated codes (suffixes “00001” - “99999”). The basic ISLI code has both “source” and “target” marked as the patent’s publication number. Each reference then receives an ISLI code with the “source” as the patent’s publication number and the “target” as the reference’s identifier (publication number, DOI, handle, ISBN-A, URN, or EDS AN). Extended metadata—including document type, identifier type, and linkage method—are written into the data-set.

Currently, China’s ISLI RA has not fully opened service code application and resolution services. ISLI application systems can establish simulation servers for code application, encoding, and resolution, with simulated codes later batch-converted to official ISLI codes once formal services become available.

7. Applications and Benefits

The constructed patent-reference linkage data-sets are reusable and shareable, enabling measurement of total, average, and extreme citation counts across different countries, regions, organizations, and patent classifications. This facilitates analysis of citation patterns and characteristics. The data-set can also be inverted to create a “mirror” data-set linking references to patents, identifying the most frequently cited journal articles, dissertations, conference papers, and monographs across disciplines. This effectively adds a patent citation database to existing citation systems for journal articles, conference proceedings, books, and scientific data.

Furthermore, the mirror data-set can be integrated into non-patent literature platforms like EDS to display patent citation data for non-patent documents (see Figure 8 [Figure 8: see original paper]), providing users with citation information. Notably, citation databases such as Web of Science and CNKI currently lack data on non-patent literature cited by patents. Integrating this patent reference data-set would significantly enrich existing citation systems, enable more accurate measurement of citation counts and academic impact, and support research on the evolution and development of fundamental research and technological innovation outcomes.

However, many early non-patent documents lack standardized identifiers like DOIs, necessitating temporary use of EDS AN numbers. Since non-EDS users cannot resolve these AN numbers, the usability of the linkage data-set is somewhat limited. As authoritative identifiers like DOI continue to be retroactively assigned, reliance on EDS AN numbers will decrease, progressively enhancing the data-set's usability.

References

- [1] Chen K, Xu F, Cheng R. Research progress on non-patent citation analysis[J]. *Library and Information Service*, 2015, 59(5): 137-144.
- [2] Li S, Guo J, Sun M. Patent search techniques and methods: Citation search using Okedata intellectual property resources as an example[J]. *Henan Science and Technology*, 2018(6): 59-61.
- [3] Narin F, Noma E. Is technology becoming science?[J]. *Scientometrics*, 1985, 7(3-6): 369-381.
- [4] Guan J, He Y. Patent-bibliometric analysis on the Chinese science-technology linkages[J]. *Scientometrics*, 2007, 72(3): 403-425.
- [5] Han F. Research on the “science-technology relationship” and technology evolution trajectory based on patent citations[D]. Beijing: Beijing University of Posts and Telecommunications, 2017.
- [6] Meyer M S. Patent citation analysis in a novel field of technology: an exploration of nano-science and nano-technology[J]. *Scientometrics*, 2001, 51(1): 163-183.
- [7] ISO 17316:2015 Information and documentation—International standard link identifier (ISLI)[EB/OL]. [2021-09-16]. <https://www.iso.org/standard/59560.html>.
- [8] Cai X. Association: The concept and value of ISLI international standard[J]. *Publishing Research*, 2015(7): 39-41.
- [9] Zhu J, Li X, Ren X, et al. International standard link identifier and its application in libraries[J]. *Library Development*, 2020(2): 56-61, 69.
- [10] Guide for using ISLI (information content industry)[EB/OL]. [2021-09-16]. https://www.isli-international.org/download/Guide_{{{for}}}{{{using}}}{ISLI}.pdf.

Author Contributions

Zhu Jiang: Conceived research ideas and wrote the paper;

Zhang Di: Revised and supplemented the paper;

Liu Chunjiang: Developed patent big data service platform, analyzed patent reference types and examples;

Zhou Qi: Designed and implemented resource discovery system linkage solutions.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv –Machine translation. Verify with original.