

## Digitization and Instant Reference System for Dharmaśāstric Knowledge Tradition

Arooshi Nigam, M.Phil. Research Scholar

Subhash Chandra, M.Phil., Ph.D.

Email: [arooshinigam@gmail.com](mailto:arooshinigam@gmail.com) and [schandra@sanskrit.du.ac.in](mailto:schandra@sanskrit.du.ac.in)

### Abstract

The tradition of *Dharmaśāstra* (DS) carries extensive history and contains the treatises of Ancient Indian Social institutions. Dharma is a key concept of the text and it signifies the behavior that is considered to be in accordance with the 'right way of living' and 'path of righteousness'. DS covers a wide range of topics and is particularly divided into three major areas as *ācāra* (Customs), *vyavahāra* (stated legal procedures) and *prāyaścita* (penance/atonement). The texts of the DS are primarily written in the Sanskrit language and later translated into various languages. In the era of Information Technology, everyone wants to obtain and share information online only. DS texts are not only studied by Sanskrit scholars but also referred by historians, sociologists, political scientists, economists, law enthusiasts and linguists worldwide. As of yet, there is quite a lack of development done in this area where a person can search the information regarding the DS texts online for further research or in-depth studies. Therefore, the major objective of the paper is to digitize and develop an instant referencing system for DS texts. Therefore, for the above-stated purpose, the complete data of the *Manusmṛiti* (MS) has been collected and digitalized in the Devanagari script in Unicode. Therefore, a web-based Information Extraction system is being developed. The proposed system will be able to take input in various scripts and generate the desired output in a similar way. The information can be searched by typing a 'word' or selecting the options from the listed DS concept menu. As a result, the complete reference of the searched words/ concepts inclusive of its bilingual (Hindi and English) translation and transliteration will be easily accessible to the user. The proposed system will be available online for public use at no cost. This system may play a very important role in teaching, learning and research because of the lack of existing effective online tools.

**Keywords:** *Dharmaśāstra* (DS), Indian Knowledge Tradition, Information Extraction, Data Mining, Online Indexing, *Manusmṛiti* (MS)

### 1. Introduction

DS is a collection of many theological texts, and it contains the treatises of Hinduism on *Dharma* (duties). *Dharma* is of primary importance in Indian philosophy, religion, social issues and community laws. Ancient Indian Social institutions represent a distinct tradition of civilization formation, society development and community living. The fundamental meaning of *Dharma* in DS is diverse and it

includes accepted norms of behavior, procedures within a ritual, moral actions, righteousness and ethical attitudes, civil and criminal law, legal procedures and penance or punishment, and guidelines for proper and productive living. Olivelle (2004) has described the concept of *Dharma* which includes various social institutions discussed in DS literature such as marriage system, inheritance property, adoption rules and regulations, work contracts rules, judiciary system and process in case of disputes within the community and society, as well as personal choices such as food preferences, suspension of study and sexual demeanors etc.

The word *dharma* is derived from the Sanskrit verb root “*dhr*” which means to hold or support (Kane,1968). It can be considered analogous with the “right way of living” or “the path of righteousness” (Chambers, 1999). It also contains the widespread ideas for proper behavior leading to productive living, rights, duties, character, customs, conduct, virtues, laws and behavior considered appropriate, correct or morally upright. *Smṛti* has humanly authored written scriptures that contain the collected traditions. The *smṛtis* is an amalgamation of sutras (prose) and DS (poetry). *Smṛtis* can be considered as an ancient framework of socio-cultural law. The code of Manu, the code of *yājñavalkya* and the code of *Nārada* are the principal *smṛtis*. All these *smṛtis* contain special sections on the social laws, consumer laws, tribal community laws and related legal procedures. There are huge *tīkās* (commentaries) and *nibandhas* (digests) which are written on the basis of *smṛtis*. The major focus of these is to interpret the law as laid down in the *smṛtis*. The commentaries and digests have systemized, consolidated and analyzed the *smṛtis*, thereby giving an entire body of law. These are of particular importance and are essential as they help in understanding the original sources in a more refined manner. Famous *tīkās* are *smṛticandrikā*, *ratnākara* and *caturvargacintāmaṇi*, etc. When humans began to live in groups, it led to the development of the concept of civilization and community living. It was then but natural that they should conform to certain patterns of human behavior for peace and harmony. This gave birth to the norms or ways of living a more structured life. This concept in the ancient DS texts was coined as *ācāra* (Kaundinyayayn, 2014).

MS is a primary text of the DS, it is an ancient legal text divided into 12 chapters (*adhyaya*). The text is composed in metric verses, in the form of a dialogue between an exalted teacher and disciples who are eager to learn about the various aspects of *dharma*. The first 58 verses are attributed to Manu, while the remaining more than two thousand verses are attributed to his student Bhrigu (Olivelle, 2005). It contains special sections on the social laws, consumer laws, tribal community laws and related legal procedures. The first chapter discusses the genesis of the universe, the origination of ten Manus, the fourfold system of *varṇa* (Sarda et al, 2018) and *āśrama*, four types of *yugas*, division of time and centuries. The second chapter explains about features and sources of *dharma*, sixteen *saṃskāra*, procedure and methods of observing vows and performing ceremonies, student- guru bond, results of *yajña*, the importance of *ācārya*, *upādhyāya*, *guru*, *ṛtvika*, the concept of *vrātya*, *brahmacharya vidhi* and so on. Third, the fourth and fifth chapter talks about the *gṛhastha āśrama* and various duties/responsibilities attached to it, householder-ship and its objectives, selection of the clan, worshipping of women in the society, hospitality towards guests, types of marriages, pacification of the deceased, livelihood aspect of life, good and bad conduct, the concept of edible and inedible. Sixth chapter deals with the *vānaprastha vidhi parivrājaka-dharma*, *saṃnyāsa* and *mokṣa*. The seventh

chapter talks about the ruler, state-ship and kingdom, duties of the king and its subjects. The eighth chapter deals with *vyavahāra darśana* i.e. monetary, financial and fiscal department, revenue and expenditure, commercial activities, duties of witnesses, etc. The ninth chapter discusses the duties of the men and women, division of property, twelve types of sons, elimination of criminals and treatment of *vaiśyas* and Shudras. The tenth chapter liberally discusses inter-caste marriages and one's duties during emergency times. The eleventh and twelfth chapter describes the penance that a man must observe to repent the sins he had committed either knowingly or unknowingly. It talks about both direct and indirect penance, means of salvation, the concept of sins, deeds and misdeeds, donations and unification with the supreme soul. Therefore, the major concepts of MS are Creation of the world, Sources of dharma, the dharma of the four social classes (Chaubey, 2005) or the fourfold system of *Varnas* and *āśrama*, law of karma, rebirth and final liberation, Virtues and outcasts, Behavior, moral code and conduct, Sources of Law, Rights of women, *puruṣārtha*, *saṃskāra*, Rights and duties of every citizen, Statecraft, Rules of War, Ruler and Kingship.

Figure 1: User Interface

## 2. Scope of the Paper

DS is reckoned as one of the major subject streams in Sanskrit and political science departments in India and abroad. MS as a text along with its commentaries is precisely taught as a component subject at postgraduate level at various departments in all major universities worldwide having Sanskrit or Indic languages department. In today's era of globalization and information technology, where the entire world is connected by a click of a button, the world news is generated, shared and received through web consortiums, yet, any instant information retrieval system based on DS where a person

can find desirable output for their queries related to this specific knowledge field is not available. The major research problem of this work is that there is not any online system available where DS texts can be accessed online and the conceptual searching from the DS key texts can be done for the complete reference.

The system is being discussed here is a Web-based Search Mechanism and information/data Extraction Mechanism for MS. This is an automated or more precisely, a digitalized system of accepting input and generating output; therefore, gathering a very specific yet in-depth knowledge of the varied concepts of MS that any user/researcher/student wishes to study. The major objective of this paper is to digitize the DS literature and development of automatic information search system for instant reference with its English and Hindi translations. The system generates output and displays result as per the client's input query.

### 3. Data Collection and Digitization

The information extraction System is an online search system for MS, developed by the Computational linguistics R&D, department of the Sanskrit University of Delhi. The user can search any word available in MS and get a complete reference of that word. Verses of MS are stored in a digital database with its reference in UTF-8 format in Devanagari script. In layman's language, data or the collection of MS shlokas are extracted, proofread and digitally stored. The entire programming of the developed search mechanism is based on the process of online indexing of the database (Chandra and Anju, 2018).

#### 3.1 Development of Computational Platform and Techniques

MS Search or Instant Reference System for MS is a cohesive mechanism as it works with the help of many small digital components. The major components are User Interface, Preprocessor, Information Extractor, Information Generator, Meaning Generator, Concepts & Script Validator and Output Generator. The computational environment for developing this search mechanism was created through building various different databases and modules. A digital database which consists a total of 2688 verses of MS is created. A concept list extracted from the MS is also created. All MS verses are translated into English and Hindi language and also a module is developed to transliterate in Roman. A Script Validator Module checks or validates the script of the input; it validates the script of the user's input and differentiates as to whether the given query is in Devanagari (वर्ण) or IAST (used for roman input such as English: *varna*) to produce the result in the same format. Information Extractor module works as an index for the MS database and further acts as an accurate referencing of the user's input for the output generation. Finally, the above modules and databases are preprocessed simultaneously and thus creates an interactive search mechanism to generate the desired output for the given query.

#### 3.2 Tools and Techniques

MS Search or Instant Reference System is a web-based system. Therefore, to develop it, the methods of information extraction of web technology and computational linguistics for search have been used. Data mining techniques are used for searching (Chandra and Anju, 2017). Databases and Text files are used for storing the data digitally any web-based system has two major parts: Front-End

and Back-End. Front-end is developed using HTML (Hyper Text Markup Language) with CSS (Cascading Style Sheets) and JS (Java script) used for page decoration and beautification of the User Interface. The back-end contains a lot of programming language, databases and servers. For this, Python programming language, Text files and Flask as a server have been used. Text editors are used to edit the extracted data.

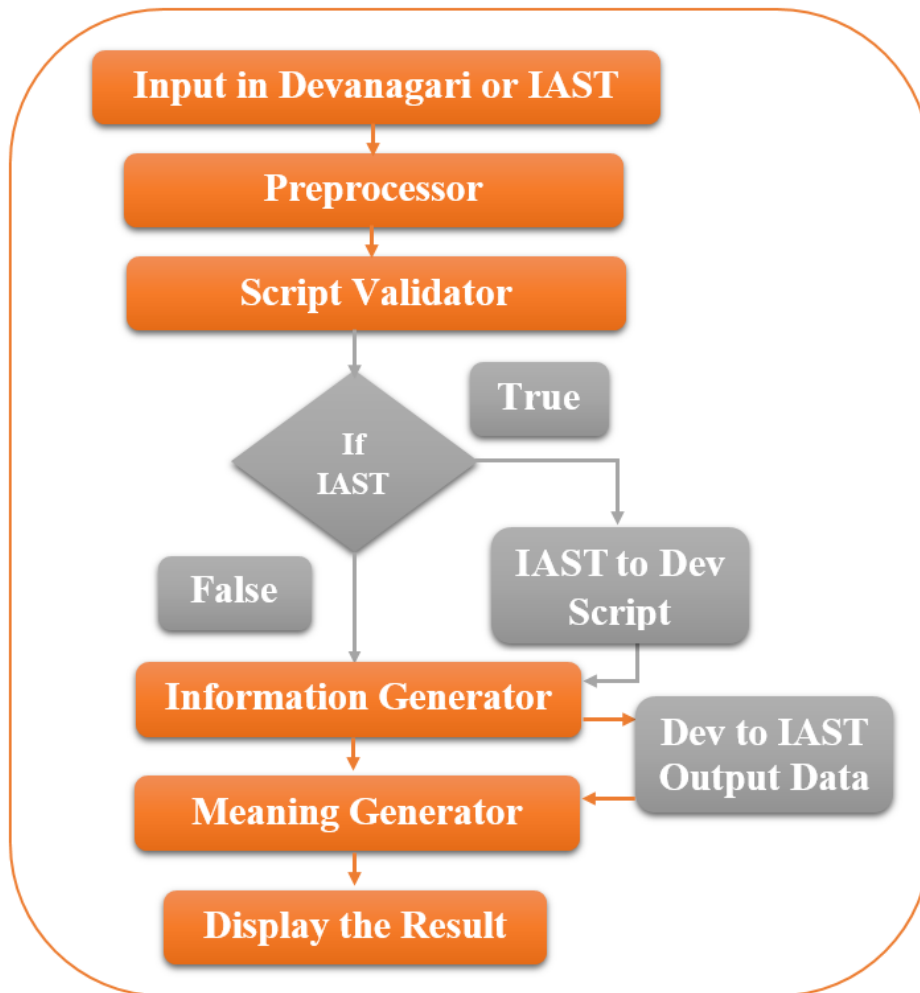


Figure 2: Flowchart of the System

### 3.3 Methodology

The developed system is an input-output generating system. It takes input from the user and generates the corresponding output. The user can give the input in either of the languages; English (roman) or Hindi (Devanagari) based upon his or her comfort level. The system has been designed in an user-friendly manner, thus, the input in the system can be given in two ways; that is either by typing the query in the text box given in the user interface (figure 1) or select an option from the pre-constructed drop-down menu. Once the input is given, a lot of pre-designed functions work simultaneously to give the output. The preprocessor initially runs the query at back end syncing it with the digital information indexer. The script validator checks the input language, concept indexer matches the tags of the respective verses with the given input query, then the following query is searched one by one from different databases and corresponding result is generated. The generated

result is formatted according to the user's query input and then displayed on the clients end. The final formatted display of the result. The working of the above explained methodology can be understood with the help of the following flowchart representation.

## 5. Major Features of the Developed System

This web-based developed system consists of a variety of features, it accepts the input query in two languages. Hence this system provides for a multiscript search in Devanagari and Roman. It further has the scope for creating multilingual searches such as Punjabi, Bangla, Oriya, etc. Since the system is digitalized, it is widely accessible. Precise conceptual information retrieval is the major highlight of this system. Information retrieval is quick and error-free. The concept of any text cannot be mined through instant indexing. So a concept list is generated, and texts are tagged with the concept. Most verses clearly denote the concept of *varṇa* system but the word "*varṇa*" doesn't appear directly in verses. The general index does not work to mine these verses for the concept "*varṇa*". Thus, these type of verses can be only mined through the conceptual searching. This is the special feature of this system, and it can be searched via this system.

This system provides for interactive data search, which gives the system a very user-friendly approach. Various options for search such as Keyword, concept and Phrase searching have been developed. Online indexing (Mani, 2010; Khandoliyan, 2012), bi-lingual translations and transliterations have also been made possible in this system. The entire system is available for public access over the web.

## 6. Proposed Result and Future Directions of the Research

MS search mechanism is a very useful system for researchers for immediate reference search. Users are able to get the respective shlokas along with its reference number, its bilingual transliterations (English & Hindi) and as well as bilingual translations of the respective shlokas. Concept based division has been also done for the ease of usability for the user. Currently this system is under development. In future, it is planned to digitize all major DS texts such as *Nārāda smṛti*, *Yājñavalakya smṛti*, *Arthaśāstra*, etc. The input/output methods of the system will be made multilingual (Punjabi, Sanskrit, Bangla, Telugu, Tamil etc.) in near future too. This web equipped system will hopefully be a major contribution in the Government of India run Digital India scheme and it will be a major education tool in helping both teachers and the students.

---

---

### References

1. Chambers, *Chambers Dictionary of Etymology*, Chambers, 1999.
2. Chandra, S., Anju. Puranic Search: An Instant Search System for Puranas. *Language in India*, 17(5), 2017.
3. Chandra, S., Anju. sāmkhya-yoga darśana paribhāṣā database evaṃ online khoja. *Research Review International Journal of Multidisciplinary*, 3(11), 890-894, 2018.
4. Chaubey, Shubhra. vedomṃ meṃ dharma kī avadhāraṇā. Ph.D dissertation, Dr. Ram Manohar Lohiya Awadh University, Faizabad, 2005.

---

---

**Language in India** [www.languageinindia.com](http://www.languageinindia.com) ISSN 1930-2940 22:02 February 2022

Arooshi Nigam, M.Phil. Research Scholar and Subhash Chandra, M.Phil., Ph.D.

Digitization and Instant Reference System for Dharmasāstric Knowledge Tradition

5. Kane, P.V. *History of Dharmashastras*. Bhandarkar Oriental Research Institute. Pune, 1968.
6. Kaundinyayayn, Shivraj Acharya. *Manusmriti*. Chaukhamba Vidyabhawan, 2014.
7. Khandoliyan, B. R., Pandey, R. K., Tiwari, A., & Jha, G. N. Text encoding and search for Āyurvedic texts: An interconnected lexical database. *Adaptation of Language Resources and Tools for Processing Cultural Heritage Objects*, 2, 2012.
8. Mani, Diwakar. RDBMS Based Lexical Resource for Indian Heritage: The Case of Mahābhārata. *In International Sanskrit Computational Linguistics Symposium*, pp. 137-149. Springer, Berlin, Heidelberg, 2010.
9. Olivelle, Patrick. *Manu's code of Law: A critical Edition*. Oxford University Press. USA, 2005.
10. Olivelle, Patrick. *The Law Code of Manu*. Oxford University Press, 2004
11. Sarda, M., B. Deshpande, S. Deo & R. Karanjkar. 2018. A comparative study on Maslow's theory and Indian Ashrama system." *International Journal of Innovative Technology and Exploring Engineering* 8(2): 48-50.



**Ms. Arooshi Nigam**

M.Phil. Research Scholar

Department of Sanskrit, Faculty of Arts, University of Delhi

Delhi-110007, India

[arooshinigam@gmail.com](mailto:arooshinigam@gmail.com)



**Dr. Subhash Chandra, M. Phil., Ph.D.**

Assistant Professor

Department of Sanskrit, Faculty of Arts, University of Delhi

Delhi-110007, India

[schandra@sanskrit.du.ac.in](mailto:schandra@sanskrit.du.ac.in)

### Article Publication

**Published Online: 28 February 2022**

**Date of Acceptance: 23 February 2022**

### \*Corresponding Author

👤 **Dr. Subhash Chandra, Email: [schandra@sanskrit.du.ac.in](mailto:schandra@sanskrit.du.ac.in)**