

Mining the Treasure of Palm Leaf Manuscripts through Information Retrieval Techniques

Bhupendra Singh¹, Neelu Jyoti Ahuja²
¹University of Petroleum and Energy Studies
Dehradun
bhupendra.bisht.59@gmail.com
²University of Petroleum and Energy Studies
Dehradun
Neelu@ddn.upes.ac.in



ABSTRACT: *Palm Leaf Manuscripts are historical heritage of India. They store within themselves the unexplored treasures of knowledge. Though the age of these leaves is about 300yrs, the tradition of guru-shisya relationship has made their life much beyond. The knowledge in them covers all aspects of life. Be it religious beliefs, literature, science, mathematics, or any other. However, due to discontinuation of practise of copying their content on fresh leaves, they now possess a fragile life which needs to be preserved at the earliest. The modern means of digitization can help in their preservation. With computer technology penetrating every aspect of life, we can exploit information retrieval algorithms to help build a system which can dig into the ocean of knowledge from these manuscripts. This paper is an attempt to popularize information retrieval from palm leaf manuscripts among computer scientists to make available the guidance of our age old heritage in shaping our future.*

Keywords: Palm Leaf Manuscript, Information Retrieval, Ancient Indian Science, Algorithm, Computer Intelligence

Received: 8 April 2019, Revised 15 July 2019, Accepted 1 August 2019

© 2019 DLINE. All Rights Reserved

DOI: 10.6025/jcl/2019/10/4/126-133

1. Introduction

INDIA is known as one of the home, of world's ancient knowledge. Many of the modern day discoveries can be linked to the foundation which our ancestral scholars have laid in past. The transmission of knowledge from ancient historical time was initially through oral mode and much later the writing system was developed. History of writing in India can be dated back to 3rd millennium BCE in the form of Indus Script by Harappa people (non-deciphered till today), and Ashokan edict in the form of Brahmi script. But they all are on stone as carvings.

The paper for writing which we use today, is credited to China, however the Chinese kept this knowledge of producing paper as close guarded secret for many centuries. Later, it reached Arabs from a captured Chinese prisoner. Before this invention ancient sages and scholars had been using leaf of Palm Tree for writing since 5th century BCE, reason being its predicted life of 200-300 years. One of the oldest surviving palm manuscripts, a Sanskrit text Shaivism, dates back to 9th century CE. It was discovered in Nepal and currently preserved in the library of Cambridge University.

One can look into predicting the possible future only after knowing the past. Indeed, today all the inventions and discoveries are happening based on our past acquired knowledge. Our phone rings and connects to another person sitting anywhere in the world based on the wireless waves which are transmitted from one place to another through the satellites revolving around the earth all the time, which came to their position after escaping gravitational pull and this force was studied and interpreted many centuries ago by Newton, ancient Indian scholar Bhaskaracharya (BhaskarII) [1] and many others. In fact, each of our minute activities in our daily life is based on the knowledge passed on to us by our ancestors. “Food is essential for our survival” is taught by nature, but the plants eaten by people at some point in history, goes through series of transformation, meeting and mating with other transformed materials all originating in the universe, before finally reaching to our stomach. But as our memory is limited we carry with us information belonging to our life events only. Information which is vital to our survival is embedded in our genes, but the societal and personnel significant work has to be remembered and passed on to next generation either vocally or in the form of writings and pictures. The Hindu holy texts, the four Vedas were first passed over orally, preserved with precision, up to the end of first millennium BCE. The ancient scholars (Rishis) would listen, remember and chant the hymns to their disciples who would repeat and the same cycle continued.

A language is what spoken by people and a script is a diagrammatic representation of thoughts. With the advent of writing system in India around 3rd century BCE in the form of Brahmi script, it was possible to capture more and more significant events and observations. In the later centuries, Palm leaf manuscripts became the most common form of writing in India and also in South Asian countries like Nepal, Sri Lanka, and Indonesia. Since ancient times, these manuscripts have been the source of knowledge for religion, understanding universe (its nature, laws, and language) and help us progress our life to be better, as a guiding Guru.

2. Palm Leaf preparation for Writing

The palm leaves are collected from petioles of palm tree and then cut into sizes as required. They are then boiled in water and later on after drying, placed between two wood planks of the same size as the leaves. After two days the leaves are taken out, polished and then dried again for two more days. Finally, the leaf is ready. But the writing on the leaf is not as simple as writing on paper with pen. The writing on the palm leaves is carried out by the practiced professional called Lipikaras, by scribing with the help of metal Stylus.

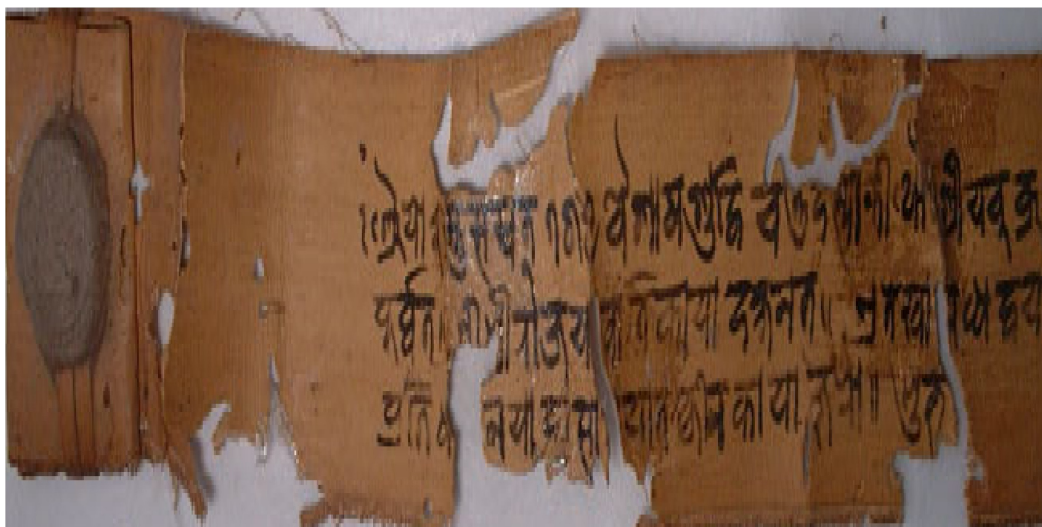


Figure 1. Palm Leaf Manuscript

IC: Wikipedia.org (https://commons.wikimedia.org/wiki/File%3ANandinagari_Manuscript.jpg)

3. History stored on Palm Leaf Manuscripts

The amount of palm leaf manuscripts available is of the order of millions. Some insight of the significance of Palm Leaf Manuscripts is provided by Anupam Sah[2]. Here we are highlighting only a few areas of work.

4. Literature

Besides religion and war, in literature, one of the greatest Sanskrit poets ever, Kavi Kalidasa, of 4-5th century, in his poem *Meghdootam*, with his vivid imagination, has inspired poets of coming centuries. In his another masterpiece, a play *Abhijñānaśākuntalam*, with his imaginative power, he brings the character of *Shakuntala* to life. The latter is the mother of Bharata, who gave the original name to a country, today known as India. As Pandit Jawaharlal Nehru notes, Kalidasa's work excludes life's misery and colors life with the beauty of nature. Unfortunately, only 5% of his work survived upto this century.

Kamasutra is well known for its explicitly erotic acts and teachings. Numerous manuscripts with titles *Smaradipika*, *Pancasayaka*, *Ratimanjari* can still be found in Orissa State Museum under the same discipline.

5. Mathematics

Dr. K V Sarma, a well-known manuscriptologist, rightly notes that, it would seem that most of the work had been done on epics, literature, philosophy and religion, which may overshadow the impressive quality of record of works in science, despite the fact that a great deal of work has been lost or destroyed over many centuries. Undoubtedly, there is a loss of important work as ancient authors Crake and Sarnagadhara refer to much earlier texts which are no longer available. Below we have briefly described few great ancient mathematical texts and scholars.

Yuktidipika (Lamp of Astronomical Rationale): A 16th century commentary by *Sankara* on *Tantrasangraha* of Nilakantha Somayaji (born 1443 CE) supplies detail rationales of the theory of numbers, division, square and square root, summations of natural numbers, squares, cubes, summations of summations, rules relating to triangle, circle and many others.

Yuktibhasa: A major treatise on mathematics and astronomy by Jyesthadeva (1503CE) in the Kerala School of Mathematics include work on infinite series expression of a function, power series, Taylor series, trigonometric series of sine, cosine, tangent and other functions. It is believed to be the foundation of calculus and predates those of James Gregory and Newton by many centuries [3].

Brahmasphutasiddhanta: The work of Brahmagupta (628C) contains the rules for manipulating both negative and positive numbers and a method for computing square root. It contained the first clear description of the quadratic formula which we use today.

Surya Sidhanta: A 6th century CE text gives estimates of the diameters of a number of planets, for example Mercury, Saturn with an error of less than 1%.

Chandahsastra: was written by Pingala, a 3rd-2nd century BCE scholar who worked extensively in combinatorics and also developed a binary system for describing prosody. He used binary number in the form of long and short syllables.

Aryabhata: A 5th century scholar who believed in geocentric model of the solar system. However, he presented many ideas which laid the foundation of modern astronomy and mathematics. He asserted that moon shines by reflecting light of Sun and correctly explained the causes of Solar and Lunar eclipse, and stated that earth rotates about its axis. He calculated value of Pi (δ), and the length of the year as 365 days 6 hours 12 mins 30 sec only 200sec longer to modern accepted values. Aryabhata's using of sine (*jya*), cosine (*kojya*) and inverse sine (*olkram jya*) influenced the birth of trigonometry.

Bhaskara II: His work *Siddhanta Siromani* is divided into four parts *Lilavati* (Arithmetic), *Bijaganita* (Algebra), *Grahaganita* (mathematics of planets) and *Goladhyaya* (Sphere) which are themselves considered as four different individual works. His work on calculus predates Newton and Leibniz by over 500 years.

6. Medicine

The geography of India has vast wealth of flora and fauna, that may be the reason for its enormous knowledge of medicine in the form of Ayurveda. The foundation of Ayurveda is based on two ancient texts that have still survived.

Charak Samhita: The 2nd century text containing eight books and one hundred twenty chapters. Mentioning medicinal uses

of 177 substances of animal origin, 341 medicinal plants, and 64 animals. It also includes section on the importance of diet, hygiene, causes and prevention of diseases and others.

Sushruta Samhita: A 2nd century BCE text on medicine and surgery. There is only a small subset of manuscripts that are printed, and still a critical edition of it is yet to be prepared. The text talks about the efforts for prevention of diseases. It also discusses surgical procedures for plastic, rhinoplastic, lithotomic and obstetrical procedures.

In a famous ‘rhinoplasty’ operation of Pune in March 1793, an Indian surgeon used skin from the foreheads of an injured soldier to repair his noses. Two senior British surgeons witnessed this operation and prepared its detailed descriptions with diagrams. The publication in Europe in 1816 of their account later gave birth to modern plastic surgery [4].

In a paper by Bhattacharya et. al. [5], he has listed some great manuscripts from Orrisa State Museum in the area of Ayurveda. For e.g *Cikitasamanjuri* belonging to 7-9th century CE while *Vesaja Ratnavali* deals with surgery. There are lots of other manuscripts as well which talk of Syrups, avian Medicine, Pediatric Medicine, Chronic diseases, blood and septum based pathology etc. In one rear treatise on cosmetology, Kalahandi Naresa, treatment for alopecia is described, a diseases incurable in modern science.

In Ayurvedic literature, bulk still remains in the form of unpublished manuscripts. A 12th century alchemist, *Vagbhata*, in his *Rasaratnasamuccaya*(ch4) talks about the medicinal aspects of gems and the curative properties of medicines using their ashes.

7. Collection of PLM in India

Orrisa State Museum: A vast collection of around 50,000 manuscripts from a range of topics, one such is the rare collection of an Ayurveda manuscript named *Cikitsamanjuri* belongs to 7-9th century CE and other *Vesaja Ratnavali* which deals with surgery. The Museum has also developed an online system from where one can purchase digital images of listed Manuscripts.

7.1.1. Government Oriental Manuscripts Library and Research Centre Chennai: The center started in year 1869 CE and has enormous collections of ancient knowledge with a total of 50,180 palm leaf manuscripts. For example, A manuscript of an author who lived before kalidasa period, around 6th century, A manuscript on Architecture, *Amsumad-bheda*, another manuscript *Kapphanabhyudaya* dealing with the life of king *Kapphana* written around 9th century CE. The collection also contains some non-deciphered manuscripts like *Gangavamsanucarita*, which is a poem giving the history of princes, who ruled the Kalinga kingdom (*Cuttack*). The institute also publishes the unpublished manuscripts annually.

7.1.2. Oriental Research Institute and Manuscripts Library, Chennai: It is an academic department which comes under the University of Kerala. The institute has around 50000 collection of PLM covering almost every field like Philosophy, Nyaya, Kavya, Tantra, Jyotisha, Shilpa, Vedanta, Astronomy, Astrology, Medicine, *Mantra Vastuvidya* (Traditional architecture) and many others.

7.1.3. The Asiatic Society, Bengal: The Asiatic Society was established in the year of 1794. The collections of manuscripts cover from the period 7th century to the 19th century. Very few of rarest collections of the museum are *Kiranavali*, Horoscope of a Muslim of the Mughal Court (1640CE.), a text on Buddhist Nyaya, A Deed of Mortgage (1639), Vajrayana text (11th c.) and others. The manuscript of *Kubjikamatam* is of the 7th century CE and the manuscript of Rigveda *Padapatha*, copied in 1362 CE, seems to be the oldest manuscript of Rigveda.

7.1.4. Degradation of PLM over the Time

The palm leaf manuscripts have a natural life of 250-300 years and then historically, the scholars used to copy their information on fresh leaves and this process continued. In the last 200 years, this chain has been broken and now we are losing these historical heritages very fast due to degradation. The threats to manuscripts are numerous, some being natural like dampness, fungus, discoloration, seepage of ink, smearing along cracks, some being the damage because of ants, rats, cockroaches, termites, others are human’s mishandling and sometimes natural disasters like floods and earthquakes.

The ASA archives in Nepal houses more than 6,700 manuscripts from the medical, tantric, astrology, Buddhism, Vedic, Purana

and many others. Though they have digitized almost all of their collection but the damage already done is not irreversible. Out of the 400 rolls of *tamsuk* (land grant documents), 151 have been chewed away by rats. Many manuscripts underwent vertical cracks, tears, folds because of the rolls being pressed down over the years [6].

Traditionally the preservation of the manuscripts has been done through indigenous methods as per Sahoo et. al. [7] like

1. Dried powdered leaves of *Aswagandha* in small packets kept with the manuscripts and covered in clothes to repel insect attack.
2. Coatings of lemon-grass oil to strengthen the leaves of manuscripts and also destroy the growth of micro-organisms.
3. Powdered *Ajwain* is also used as an insect killer and fungicide.

But today's technology requires the use of Digitization for their life long preservation, of the precious knowledge present on them.

7.1.5. Treasures still in Dark

As per Prof. David Pingree, there is estimate of about three million manuscripts still existing. Regrettably, only few of them have been subjected to modern analysis and gradually the whole ensemble is decaying. And since there are only few scholars trained to read and understand these texts, it is of utmost risk that most of these manuscripts may disappear before anyone is able to analyse and describe their content correctly.

According to UNESCO, more than a hundred thousand palm-leaf manuscripts are still unpublished on various aspects of traditional Indian knowledge in Tamil itself. One example is tantra theology manuscript, *Paramesvaratantra* of year 828 CE in the library of University of Cambridge which is still unpublished.

Shepherds in Gilgit in Kashmir valley in 1931, discovered a large number of manuscripts opening a new chapter on Buddhist Kashmir. In India, these manuscripts are among the oldest surviving manuscripts.

Below is the list of findings of Dr. K V Sarma and his team during the extensive survey of manuscripts in 1995-97 present in Kerala

Section	Title	No. of manuscripts
I	Astronomy	2919
II	Astrology	6794
III	Medicine	1286
IV	Veterinary Science	146
V	Chemistry	166
VI	Physics	326
VII	Botany	8
VIII	Architecture	599
	Total	12244

Table 1. Manuscripts collection in Kerala and Tamil Nadu

and Tamil Nadu, in both public and private collections where 400 repositories with 150,000 manuscripts were extensively surveyed by K V Sarma and his team [8]. Sadly, only 7% of these texts are in print which is regarded as the sole contribution of India to science through ancient manuscripts.

7.1.6. Digitization Efforts

UNESCO under the program “Memory of the world” is engaged in the mission of facilitating preservation, assistance in providing universal access to heritage documents and increasing the awareness worldwide about the existence and significance of heritage documents.

7.1.7. National Mission for Manuscripts (NMM)

The National Mission for Manuscripts established in February 2003, by the Ministry of Tourism and Culture, Government of India, is a unique project committed to preserving the vast manuscript wealth distributed throughout of India. It is estimated that there are ten million manuscripts, in India. NMM is working towards fulfilling its motto, ‘conserving the past for the future’. The NMM is developing a database of all Indian manuscripts named *Kriti Sampada* which will be available through internet.

As per Ministry of state for Culture and Tourism, till 2017 2.8 lakh manuscripts have been digitized as per PIB India [9].

With its commissioning since 2003 and given the volume of manuscripts to be digitized. It is in question whether all the manuscripts would survive given their present brittle state. There is one more challenge that there are only few scholars who can decipher them. PMM has organized workshops on Manuscriptology and Palaeography to tackle this challenge. The NMM has provided 23 states with trained conservators who have discovered manuscripts in the state of Mizoram, not known before 2010.

NMM has also submitted a proposal for setting up of Digital Manuscripts Library to store Digital Manuscripts for accessibility purpose. It also has a publication program where 30 unpublished manuscripts have been published.

7.1.8. Information Retrieval System for Palm Leaf Manuscripts Repository

The Government of India and other organizations are doing commendable job of preserving and safeguarding country’s heritage and age old knowledge system through the movement of digitization. In the years to come, the agonizing problem of manuscripts degradation will be eradicated completely. However, next when it will come to mining the knowledge treasure out of these manuscripts, we would be confronted with another helpless situation. The digitization process would capture the manuscripts from present physical palm leaf to digital image form by clicking high quality pictures. All the text in a palm leaf will be available in the form of images, but on these images a simple search for any word would not be possible. Consider for example, if you are reading a PLM on astrology and want to search for your zodiac prediction, without textual format you will have to do manual search through all the leaf images involving stressful exercise of your eye.

Thus the manuscripts need to be available in textual form and not only in image form.

The next problem which we will encounter is to categorize millions of manuscripts into titles/ headings of Astrology, Astronomy, Medicine, Mathematics, Yoga, religion and so on. Doing this categorization manually would not only require massive human resource and time, but also the errors would be untraceable. We thus need, an automated system for categorization, and sub categorization.

The system should also be intelligent enough to create category/ title name itself. For example, there may be a particular disease cure in significant number of manuscripts, thus that particular disease should pop out as a sub-category by itself under medicine category. The system should also be capable of adding any new manuscript into its structure later on and arrange and rearrange itself as and when required.

The categorization will help scholars and others to search for their topic from within the different categories and sub-categories. The great king Ashoka is well known to almost all those who have gone through Indian Primary Education. Most of the story of Ashoka has been derived from a Buddha manuscript titled Ashokavadana. Thus if someone is interested in knowing about Ashoka, he would have to go through the hierarchical structure of

History -> Chronology -> Kingdoms -> Kings -> Ashoka

But here we simply want all the manuscripts which deal with the word Ashoka in their content. Thus, we would need a search engine which can retrieve all the manuscripts which deal with Ashoka.

The search engine should be intelligent in nature. It should not only pop up manuscripts on the basis of words (specifically called keywords) provided by the user but also look for similar words or synonyms. Consider for example the case that historically king Ashoka was known as Priyadasi as well. Thus the search should also include synonym names to its results. In another dimension of the problem, synonyms inclusion may fail as well. Consider again, there may be many manuscripts detailing about the battle of Kalinga, administration of Ashoka, life of people under Ashoka. All these manuscripts are closely related to the same primary subject 'Ashoka' but none may have any reference to 'Ashoka' word or its synonyms. These issues need to be dealt with the search for common linking words between different manuscripts. For example, the common link can be 'name of the city' some manuscripts are describing and, the same 'name of the city', having Ashoka or its synonym. These are the topics which come under the realm of semantic analysis.

An Intelligent Search Engine with semantic analysis is need of the hour.

Information Retrieval: All the above problems fall more or less under the purview of information retrieval. According to the definition of Raghvan [10], "Information retrieval is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)."

It works as

1. Processing all the documents and then finding all the tokens/words from them.
2. Arranging the tokens alphabetically having information about the frequency of their occurrence within the corpus and link to all the documents where they are present
3. Adding rules for associating different tokens
4. Adding semantic rules for query processing which can look for possible synonyms

After processing the query from the user it provides results with the links to different documents. The documents are presented to the user on the basis of their relevance, with the most relevant document on top of the search results.

Challenges to the proposed IR system:

Though the above discussed problems may be addressed in the proposed system but the amount or level of intelligence of the system will remain in question. For example, if someone searches with the query

Query: 'Ashoka's battle after Kalinga'

How the system would respond? Should it pop out all the manuscripts describing Ashoka's battle and the particular battle of Kalinga? How it comes up with a date which is not given in the query and also, may not be given in any particular text. The challenge of the system would enter a new dimension, if someone requires the manuscripts with the queries as

Query1: Was Ashoka cruel?

Query2: How much highest amount of money Ashoka ever had?

Answer of the first, requires boolean response in the form of yes or no, which no matter how much a group of historians/scholars could debate, never a trivial yes/no would reach. But an intelligent machine could argue with an empirical number, based on all the events where people were killed, their number, of the direct consequences of Ashoka's order and the number of events where Ashoka's decision resulted in savings of life, their number. Well again the above framework of answering is another debate in itself.

We want to leave query 2 to the reader's themselves for suggestions and see how a machine could be taught to stand equal to

human's reasoning ability.

8. Conclusion

Working towards, mining the treasure of knowledge from the palm leaf manuscripts, hordes of challenges have been outlined. Over and above the problem of preventing decay to palm leaf manuscripts, is the challenge of deciphering text, image analysis, information retrieval and search. Search is further associated with issues of meaningful and useful extraction through semantic analysis. This paper advocates the dire need for systematic research to be undertaken in this field opening up avenues for past knowledge to guide future prospects in several domains.

References

- [1] Deshmukh, P. C., et al. (2017) GTR Component of Planetary Precession, *RESONANCE*, 2017, 577-596.
- [2] Anupam Sah, Palm Leaf manuscripts of the world: material, technology and conservation, *Studies in Conservation*, 47:sup1, 15-24, 2002, DOI: 10.1179/sic.2002.47. Supplement-1.15.
- [3] Divakaran, P. P. (2010). Calculus in India: The historical and mathematical context, *Current Science*, 2010, 8-14.
- [4] 'India's Contribution to Plastic Surgery', Press Information Bureau, Government of India <http://pib.nic.in/feature/fe0299/f0602991.html>
- [5] Bhattacharya, Deepak, Historical Notes: Select Palm Leaf Manuscripts on Health Care, *Indian Journal of History of Science*, 2014, 293-297.
- [6] Naoko Takagi, Yoriko Chudo, Reiko Maeda, Conservation and Digitization of rolled palm leaf manuscripts in Nepal, Nov 14, 2005, *Asian Art*.
- [7] Sahoo, J., and B. Mohanty, Indigenous methods of preserving manuscripts: and overview, *The Orissa Historical Research Journal*, 2004, 28-32.
- [8] Sarma, Krishna V., and V. Kutumba Sastry, Science texts in Sanskrit in the manuscripts repositories of Kerala and Tamilnadu, *Rashtriya Sanskrit Sansthan*, 2002.
- [9] Digitization of Manuscripts, Press Information Bureau of India, 20-Mar-17.
- [10] Christopher D. Manning, Prabhakar Raghvan, Hinrich Schutze, *Introduction to Information Retrieval*, Cambridge University Press, 7 July 2008.