IndoWordNet Workshop, IIT Kharagpur.

The Knowledge Structure in Sanskrit kośas

Amba Kulkarni and Sivaja S. Nair

Department of Sanskrit Studies,
University of Hyderabad,
Hyderabad
apksh@uohyd.ernet.in and sivaja.s.nair@gmail.com

December 8, 2010

Route Map

- Introduction
- Amarakośa
- Vaijayantīkośa
- ConceptNet
- Knowledge structure in Sanskrit Kośas
- WordNet and Amarakośa
- ConceptNet and the Sanskrit Kośas
- Amarakośa-jñāna-jāla : Proof of concept
- Conclusion

Route Map

- Introduction
- Amarakośa
- Vaijayantīkośa
- ConceptNet
- Knowledge structure in Sanskrit Kośas
- WordNet and Amarakośa
- ConceptNet and the Sanskrit Kośas
- Amarakośa-jñāna-jāla : Proof of concept
- Conclusion

Route Map

- Introduction
- Amarakośa
- Vaijayantīkośa
- ConceptNet
- Knowledge structure in Sanskrit Kośas
- WordNet and Amarakośa
- ConceptNet and the Sanskrit Kośas
- Amarakośa-jñāna-jāla : Proof of concept
- Conclusion

Route Map

- Introduction
- Amarakośa
- Vaijayantīkośa
- ConceptNet
- Knowledge structure in Sanskrit Kośas
- WordNet and Amarakośa
- ConceptNet and the Sanskrit Kośas
- Amarakośa-jñāna-jāla : Proof of concept
- Conclusion

Route Map

- Introduction
- Amarakośa
- Vaijayantīkośa
- ConceptNet
- Knowledge structure in Sanskrit Kośas
- WordNet and Amarakośa
- ConceptNet and the Sanskrit Kośas
- Amarakośa-jñāna-jāla : Proof of concept
- Conclusion

Route Map

- Introduction
- Amarakośa
- Vaijayantīkośa
- ConceptNet
- Knowledge structure in Sanskrit Kośas
- WordNet and Amarakośa
- ConceptNet and the Sanskrit Kośas
- Amarakośa-jñāna-jāla : Proof of concept
- Conclusion

Route Map

- Introduction
- Amarakośa
- Vaijayantīkośa
- ConceptNet
- Knowledge structure in Sanskrit Kośas
- WordNet and Amarakośa
- ConceptNet and the Sanskrit Kośas
- Amarakośa-jñāna-jāla : Proof of concept
- Conclusion

Route Map

- Introduction
- Amarakośa
- Vaijayantīkośa
- ConceptNet
- Knowledge structure in Sanskrit Kośas
- WordNet and Amarakośa
- ConceptNet and the Sanskrit Kośas
- Amarakośa-jñāna-jāla : Proof of concept
- Conclusion

Route Map

- Introduction
- Amarakośa
- Vaijayantīkośa
- ConceptNet
- Knowledge structure in Sanskrit Kośas
- WordNet and Amarakośa
- ConceptNet and the Sanskrit Kośas
- Amarakośa-jñāna-jāla : Proof of concept
- Conclusion

Introduction

- In traditional circumstances, a child would receive his education largely by oral transmission, mainly through rote-learning.
- The method employed was through recitation and remembering.
- Amarakośa typically would take anywhere between 6 months to a year to memorise a list of approximately 10,000 Sanskrit words, arranged as a list of synonyms.
- The close inspection of the structure of Amarakośa gives much more insight into the way the words are organised.
- When a student memorises it, though in the beginning it appears as a linear list of words, When he starts understanding the meaning of the words, and starts using these words, the linear structure unfolds into a knowledge web with various links.

Introduction

- In traditional circumstances, a child would receive his education largely by oral transmission, mainly through rote-learning.
- The method employed was through recitation and remembering.
- Amarakośa typically would take anywhere between 6 months to a year to memorise a list of approximately 10,000 Sanskrit words, arranged as a list of synonyms.
- The close inspection of the structure of Amarakośa gives much more insight into the way the words are organised.
- When a student memorises it, though in the beginning it appears as a linear list of words, When he starts understanding the meaning of the words, and starts using these words, the linear structure unfolds into a knowledge web with various links.

Introduction

- In traditional circumstances, a child would receive his education largely by oral transmission, mainly through rote-learning.
- The method employed was through recitation and remembering.
- Amarakośa typically would take anywhere between 6 months to a year to memorise a list of approximately 10,000 Sanskrit words, arranged as a list of synonyms.
- The close inspection of the structure of Amarakośa gives much more insight into the way the words are organised.
- When a student memorises it, though in the beginning it appears as a linear list of words, When he starts understanding the meaning of the words, and starts using these words, the linear structure unfolds into a knowledge web with various links.

Introduction

- In traditional circumstances, a child would receive his education largely by oral transmission, mainly through rote-learning.
- The method employed was through recitation and remembering.
- Amarakośa typically would take anywhere between 6 months to a year to memorise a list of approximately 10,000 Sanskrit words, arranged as a list of synonyms.
- The close inspection of the structure of Amarakośa gives much more insight into the way the words are organised.
- When a student memorises it, though in the beginning it appears as a linear list of words, When he starts understanding the meaning of the words, and starts using these words, the linear structure unfolds into a knowledge web with various links.

Introduction

- In traditional circumstances, a child would receive his education largely by oral transmission, mainly through rote-learning.
- The method employed was through recitation and remembering.
- Amarakośa typically would take anywhere between 6 months to a year to memorise a list of approximately 10,000 Sanskrit words, arranged as a list of synonyms.
- The close inspection of the structure of Amarakośa gives much more insight into the way the words are organised.
- When a student memorises it, though in the beginning it appears as a linear list of words, When he starts understanding the meaning of the words, and starts using these words, the linear structure unfolds into a knowledge web with various links.

- Many other kośas such as Vaijayantīkośa, Halāyudhakośa, Trikāṇḍśeṣa, Hārāvalī, Abhidhānacintāmaṇī etc. reveal similar structure when explored.
- The modern efforts of building ConceptNet aimed at building a network around various concepts, have some parallels with the structure of these Sanskrit kośas.
- Here we present a few samples of knowledge structure from two important kośas viz. Vaijayantīkośa and Amarakośa.
- Also compare the knowledge structure involved with that of the ConceptNet and WordNet.
- We also propose a structure to represent this knowledge in e-form and a set of programs that unfold the implicit/hidden structure dynamically.

- Many other kośas such as Vaijayantīkośa, Halāyudhakośa, Trikāṇḍśeṣa, Hārāvalī, Abhidhānacintāmaṇī etc. reveal similar structure when explored.
- The modern efforts of building ConceptNet aimed at building a network around various concepts, have some parallels with the structure of these Sanskrit kośas.
- Here we present a few samples of knowledge structure from two important kośas viz. Vaijayantīkośa and Amarakośa.
- Also compare the knowledge structure involved with that of the ConceptNet and WordNet.
- We also propose a structure to represent this knowledge in e-form and a set of programs that unfold the implicit/hidden structure dynamically.

- Many other kośas such as Vaijayantīkośa, Halāyudhakośa, Trikāṇḍśeṣa, Hārāvalī, Abhidhānacintāmaṇī etc. reveal similar structure when explored.
- The modern efforts of building ConceptNet aimed at building a network around various concepts, have some parallels with the structure of these Sanskrit kośas.
- Here we present a few samples of knowledge structure from two important kośas viz. Vaijayantīkośa and Amarakośa.
- Also compare the knowledge structure involved with that of the ConceptNet and WordNet.
- We also propose a structure to represent this knowledge in e-form and a set of programs that unfold the implicit/hidden structure dynamically.

- Many other kośas such as Vaijayantīkośa, Halāyudhakośa, Trikāṇḍśeṣa, Hārāvalī, Abhidhānacintāmaṇī etc. reveal similar structure when explored.
- The modern efforts of building ConceptNet aimed at building a network around various concepts, have some parallels with the structure of these Sanskrit kośas.
- Here we present a few samples of knowledge structure from two important kośas viz. Vaijayantīkośa and Amarakośa.
- Also compare the knowledge structure involved with that of the ConceptNet and WordNet.
- We also propose a structure to represent this knowledge in e-form and a set of programs that unfold the implicit/hidden structure dynamically.

- Many other kośas such as Vaijayantīkośa, Halāyudhakośa, Trikāṇḍśeṣa, Hārāvalī, Abhidhānacintāmaṇī etc. reveal similar structure when explored.
- The modern efforts of building ConceptNet aimed at building a network around various concepts, have some parallels with the structure of these Sanskrit kośas.
- Here we present a few samples of knowledge structure from two important kośas viz. Vaijayantīkośa and Amarakośa.
- Also compare the knowledge structure involved with that of the ConceptNet and WordNet.
- We also propose a structure to represent this knowledge in e-form and a set of programs that unfold the implicit/hidden structure dynamically.

Amarakośa

- Primarily named as *Nāmaliṅgānuśāsana* (a work that deals with instructions related to the gender of nouns).
- Authored by Amarasiṃha - 4^{th} century A.D. (Oka, 1981).
- Amarakośa is the most celebrated and authoritative ancient Sanskrit thesaurus.
- It has around 60 commentaries and translations into modern Indian as well as foreign languages such as Chinese, Tibetan, French, etc..
- It consists of 1,608 verses composed in anuṣṭup meter
- Divided into 3 chapters called Kāṇḍas.

Amarakośa

- Primarily named as *Nāmaliṅgānuśāsana* (a work that deals with instructions related to the gender of nouns).
- Authored by Amarasiṃha - 4^{th} century A.D. (Oka, 1981).
- Amarakośa is the most celebrated and authoritative ancient Sanskrit thesaurus.
- It has around 60 commentaries and translations into modern Indian as well as foreign languages such as Chinese, Tibetan, French, etc..
- It consists of 1,608 verses composed in anuṣṭup meter
- Divided into 3 chapters called Kāṇḍas.

Amarakośa

- Primarily named as *Nāmaliṅgānuśāsana* (a work that deals with instructions related to the gender of nouns).
- Authored by Amarasiṃha - 4^{th} century A.D. (Oka, 1981).
- Amarakośa is the most celebrated and authoritative ancient Sanskrit thesaurus.
- It has around 60 commentaries and translations into modern Indian as well as foreign languages such as Chinese, Tibetan, French, etc..
- It consists of 1,608 verses composed in anuṣṭup meter
- Divided into 3 chapters called Kāṇḍas.

Amarakośa

- Primarily named as *Nāmaliṅgānuśāsana* (a work that deals with instructions related to the gender of nouns).
- Authored by Amarasiṃha - 4^{th} century A.D. (Oka, 1981).
- Amarakośa is the most celebrated and authoritative ancient Sanskrit thesaurus.
- It has around 60 commentaries and translations into modern Indian as well as foreign languages such as Chinese, Tibetan, French, etc..
- It consists of 1,608 verses composed in anuṣṭup meter
- Divided into 3 chapters called Kāṇḍas.

Amarakośa

- Primarily named as *Nāmaliṅgānuśāsana* (a work that deals with instructions related to the gender of nouns).
- Authored by Amarasiṃha - 4^{th} century A.D. (Oka, 1981).
- Amarakośa is the most celebrated and authoritative ancient Sanskrit thesaurus.
- It has around 60 commentaries and translations into modern Indian as well as foreign languages such as Chinese, Tibetan, French, etc..
- It consists of 1,608 verses composed in anuṣṭup meter
- Divided into 3 chapters called Kāṇḍas.

Amarakośa

- Primarily named as *Nāmaliṅgānuśāsana* (a work that deals with instructions related to the gender of nouns).
- Authored by Amarasiṃha - 4^{th} century A.D. (Oka, 1981).
- Amarakośa is the most celebrated and authoritative ancient Sanskrit thesaurus.
- It has around 60 commentaries and translations into modern Indian as well as foreign languages such as Chinese, Tibetan, French, etc..
- It consists of 1,608 verses composed in anuṣṭup meter
- Divided into 3 chapters called Kāṇḍas.

Amarakośa Classification

Prathamakāṇḍam	**Dvitīyakāṇḍam**	**Tṛtīyakāṇḍam**
Svargavargaḥ	Bhūmivargaḥ	Viśeṣyanighnavargaḥ
Vyomavargaḥ	Puravargaḥ	Saṁkīrṇavargaḥ
Digvargaḥ	Śailavargaḥ	Nānārthavargaḥ
Kālavargaḥ	Vanauṣadhivargaḥ	Avyayavargaḥ
Dhīvargaḥ	Siṃhādivargaḥ	Liṅgādisaṅgrahavargaḥ
Śabdādivargaḥ	Manuṣyavargaḥ	
Nāṭyavargaḥ	Brahmavargaḥ	
Pātālabhogivargaḥ	Kṣatriyavargaḥ	
Narakavargaḥ	Vaiśyavargaḥ	
Vārivargaḥ	Śūdravargaḥ	

Statistics

Tokens and types in each kāṇḍas :-

Kāṇḍa	tokens	types
Prathamakāṇḍam	2,465	2,300
Dvitīyakāṇḍam	5,827	5,282
Tṛtīyakāṇḍam	3,288	2,271
Total	11,580	9,853

- Uniqe types - 9,031

Vaijayantīkośa

- It is a voluminous lexicon by Yādavaprakāśa.
- There are approximately 18,000 tokens.
- Divided into two broad divisions, viz. synonym sets and polysemous words.
- There are total forty three sub-sections.

Vaijayantīkośa

- It is a voluminous lexicon by Yādavaprakāśa.
- There are approximately 18,000 tokens.
- Divided into two broad divisions, viz. synonym sets and polysemous words.
- There are total forty three sub-sections.

Vaijayantīkośa

- It is a voluminous lexicon by Yādavaprakāśa.
- There are approximately 18,000 tokens.
- Divided into two broad divisions, viz. synonym sets and polysemous words.
- There are total forty three sub-sections.

Vaijayantīkośa

- It is a voluminous lexicon by Yādavaprakāśa.
- There are approximately 18,000 tokens.
- Divided into two broad divisions, viz. synonym sets and polysemous words.
- There are total forty three sub-sections.

Synonym sets and polysemous words

- Synonym sets divided into five classes or kāṇḍa's.
  - Svarga (heaven)
  - Antarīkṣa (sky)
  - Bhūmi (earth)
  - Pātāla (nether world)
  - Sāmānya (miscellaneous)
- The polysemous words are classified into three classes based on the number of syllables they contain, viz., two, three and more than three.
- It has eight classes which are further sub-divided into two or more sub-sections called *adhyāyās*.

Synonym sets and polysemous words

- Synonym sets divided into five classes or kāṇḍa's.
  - Svarga (heaven)
  - Antarīkṣa (sky)
  - Bhūmi (earth)
  - Pātāla (nether world)
  - Sāmānya (miscellaneous)
- The polysemous words are classified into three classes based on the number of syllables they contain, viz., two, three and more than three.
- It has eight classes which are further sub-divided into two or more sub-sections called *adhyāyās*.

Synonym sets and polysemous words

- Synonym sets divided into five classes or kāṇḍa's.
  - Svarga (heaven)
  - Antarīkṣa (sky)
  - Bhūmi (earth)
  - Pātāla (nether world)
  - Sāmānya (miscellaneous)
- The polysemous words are classified into three classes based on the number of syllables they contain, viz., two, three and more than three.
- It has eight classes which are further sub-divided into two or more sub-sections called *adhyāyās*.

Synonym sets and polysemous words

- Synonym sets divided into five classes or kāṇḍa's.
  - Svarga (heaven)
  - Antarīkṣa (sky)
  - Bhūmi (earth)
  - Pātāla (nether world)
  - Sāmānya (miscellaneous)
- The polysemous words are classified into three classes based on the number of syllables they contain, viz., two, three and more than three.
- It has eight classes which are further sub-divided into two or more sub-sections called *adhyāyās*.

Synonym sets and polysemous words

- Synonym sets divided into five classes or kāṇḍa's.
  - Svarga (heaven)
  - Antarīkṣa (sky)
  - Bhūmi (earth)
  - Pātāla (nether world)
  - Sāmānya (miscellaneous)
- The polysemous words are classified into three classes based on the number of syllables they contain, viz., two, three and more than three.
- It has eight classes which are further sub-divided into two or more sub-sections called *adhyāyās*.

Synonym sets and polysemous words

- Synonym sets divided into five classes or kāṇḍa's.
  - Svarga (heaven)
  - Antarīkṣa (sky)
  - Bhūmi (earth)
  - Pātāla (nether world)
  - Sāmānya (miscellaneous)
- The polysemous words are classified into three classes based on the number of syllables they contain, viz., two, three and more than three.
- It has eight classes which are further sub-divided into two or more sub-sections called *adhyāyās*.

Synonym sets and polysemous words

- Synonym sets divided into five classes or kāṇḍa's.
  - Svarga (heaven)
  - Antarīkṣa (sky)
  - Bhūmi (earth)
  - Pātāla (nether world)
  - Sāmānya (miscellaneous)
- The polysemous words are classified into three classes based on the number of syllables they contain, viz., two, three and more than three.
- It has eight classes which are further sub-divided into two or more sub-sections called *adhyāyās*.

Synonym sets and polysemous words

- Synonym sets divided into five classes or kāṇḍa's.
  - Svarga (heaven)
  - Antarīkṣa (sky)
  - Bhūmi (earth)
  - Pātāla (nether world)
  - Sāmānya (miscellaneous)
- The polysemous words are classified into three classes based on the number of syllables they contain, viz., two, three and more than three.
- It has eight classes which are further sub-divided into two or more sub-sections called *adhyāyās*.

Classification of the synonymous part of Vaijayantīkośa

- Svarga (heaven)
  - Ādideva (supreme diety)
  - Lokapāla (guardian deities)
  - Yakṣa (semi-divine beings)
- Antarikṣaka (sky)
  - Jyoti (light)
  - Megha (cloud)
  - Khaga (bird)
  - Śabda (sound)
- Būmi (earth)
  - Deśa (place)
  - Śaila (hill)
  - Vana (forest)
  - Paśusaṅgraha (animals)
  - Manuṣya (mankind)
  - Brāhmaṇa (priest tribe)
  - Kṣatriya (military tribe)
  - Vaiśya (bussiness tribe)
  - Śūdra (mixed class)

cont....

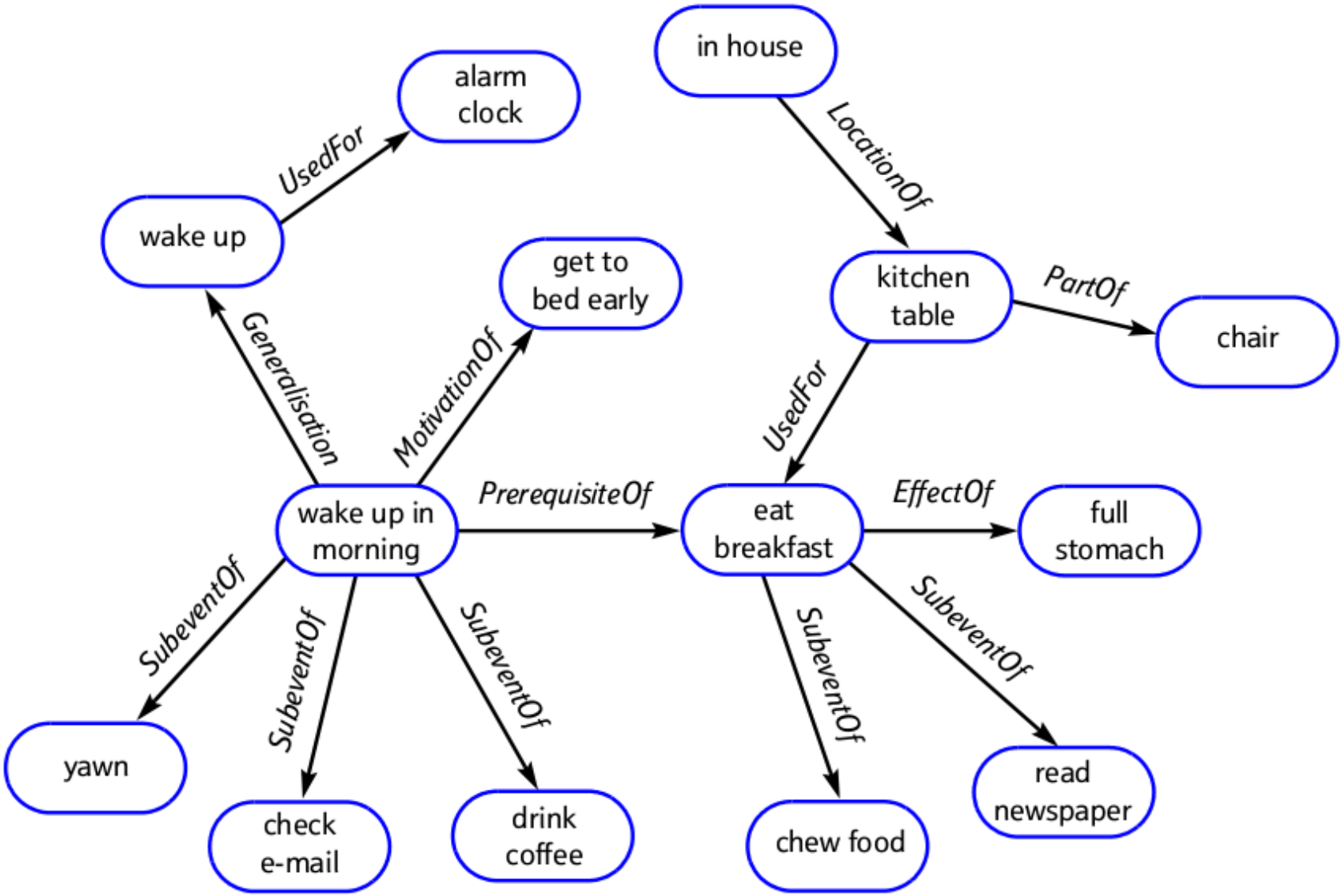
- Pātāla (nether world)
  - Sarīsṛpa (serpent)
  - Jala(water)
  - Pura (town & cities)
  - Būta(living being)
- Sāmānya (miscellaneous)
  - Gaṇa (multitude)
  - Dharmakarma (natural character)
  - Guṇa (quality)

ConceptNet

- ConceptNet is a commonsense knowledge base and natural-language-processing toolkit.
- It is a semantic network of commonsense knowledge.
- It aims to give computer an access to the commonsense knowledge - the kind of information that ordinary people know but usually leave unstated.
- ConceptNet is generated automatically from the English sentences of the Open Mind Common Sense(OMCS) corpus(Liu et. al., 2004).

Knowledge representation in ConceptNet

"wake up in the morning and drink coffee".



Knowledge structure in Sanskrit Kośas

- Except for the polysemous words (nānārthavarga), all other synsets in a class show some semantic relation, to the class it belongs to and sometimes even to the preceding or following synsets.
- These semantic relations indicate various kinds of relations.
- They may be classified as hierarchical or associative.
- Hierarchical relations
  - hypernym - hyponym
  - holonym - meronym
- Associative relations
  - association among human beings
  - association of certain objects with certain other objects.

Knowledge structure in Sanskrit Kośas

- Except for the polysemous words (nānārthavarga), all other synsets in a class show some semantic relation, to the class it belongs to and sometimes even to the preceding or following synsets.
- These semantic relations indicate various kinds of relations.
- They may be classified as hierarchical or associative.
- Hierarchical relations
  - hypernym - hyponym
  - holonym - meronym
- Associative relations
  - association among human beings
  - association of certain objects with certain other objects.

Knowledge structure in Sanskrit Kośas

- Except for the polysemous words (nānārthavarga), all other synsets in a class show some semantic relation, to the class it belongs to and sometimes even to the preceding or following synsets.
- These semantic relations indicate various kinds of relations.
- They may be classified as hierarchical or associative.
- Hierarchical relations
  - hypernym - hyponym
  - holonym - meronym
- Associative relations
  - association among human beings
  - association of certain objects with certain other objects.

Knowledge structure in Sanskrit Kośas

- Except for the polysemous words (nānārthavarga), all other synsets in a class show some semantic relation, to the class it belongs to and sometimes even to the preceding or following synsets.
- These semantic relations indicate various kinds of relations.
- They may be classified as hierarchical or associative.
- Hierarchical relations
  - hypernym - hyponym
  - holonym - meronym
- Associative relations
  - association among human beings
  - association of certain objects with certain other objects.

Knowledge structure in Sanskrit Kośas

- Except for the polysemous words (nānārthavarga), all other synsets in a class show some semantic relation, to the class it belongs to and sometimes even to the preceding or following synsets.
- These semantic relations indicate various kinds of relations.
- They may be classified as hierarchical or associative.
- Hierarchical relations
  - hypernym - hyponym
  - holonym - meronym
- Associative relations
  - association among human beings
  - association of certain objects with certain other objects.

Knowledge structure in Sanskrit Kośas

- Except for the polysemous words (nānārthavarga), all other synsets in a class show some semantic relation, to the class it belongs to and sometimes even to the preceding or following synsets.
- These semantic relations indicate various kinds of relations.
- They may be classified as hierarchical or associative.
- Hierarchical relations
  - hypernym - hyponym
  - holonym - meronym
- Associative relations
  - association among human beings
  - association of certain objects with certain other objects.

Knowledge structure in Sanskrit Kośas

- Except for the polysemous words (nānārthavarga), all other synsets in a class show some semantic relation, to the class it belongs to and sometimes even to the preceding or following synsets.
- These semantic relations indicate various kinds of relations.
- They may be classified as hierarchical or associative.
- Hierarchical relations
  - hypernym - hyponym
  - holonym - meronym
- Associative relations
  - association among human beings
  - association of certain objects with certain other objects.

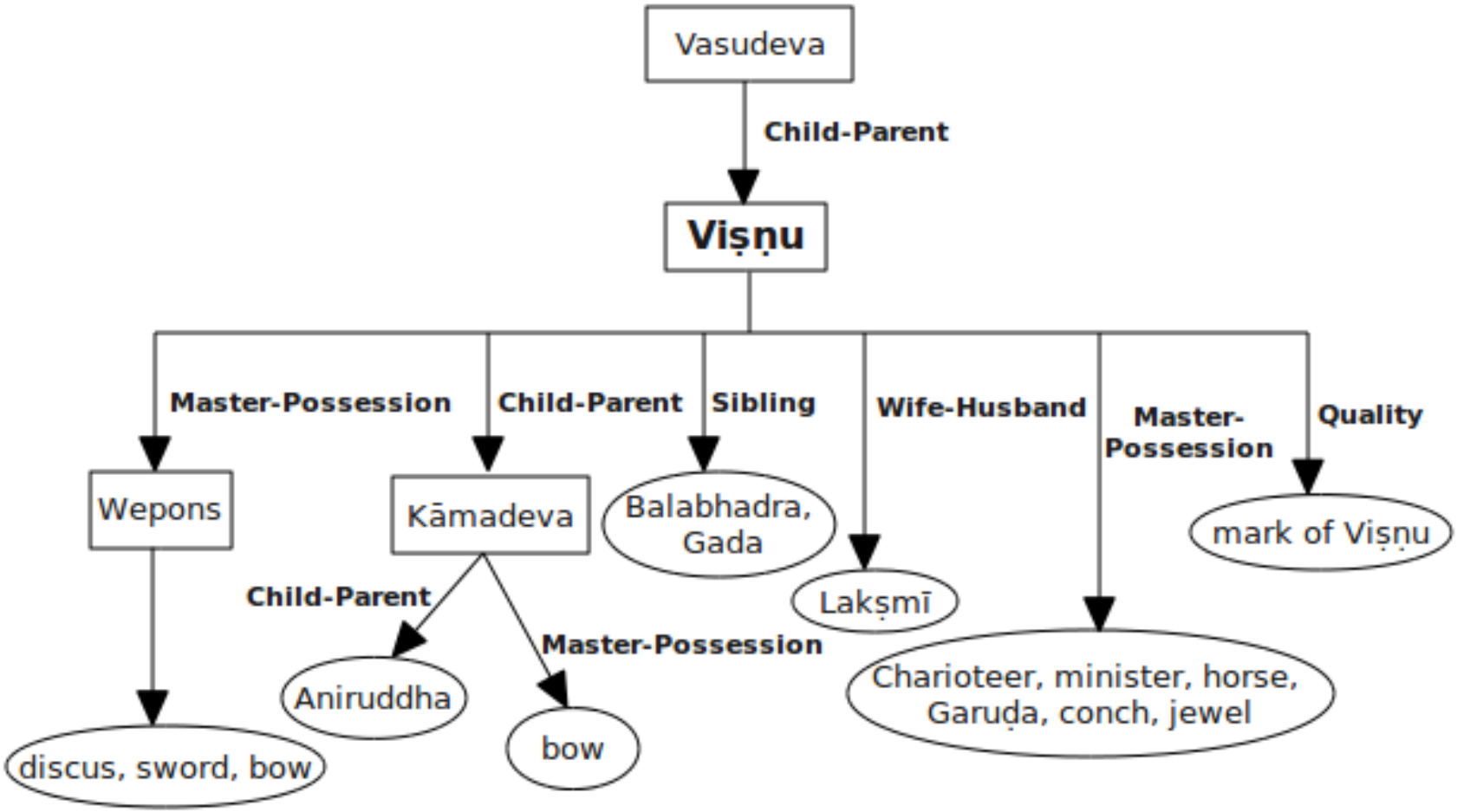
Knowledge structure in Sanskrit Kośas

- Except for the polysemous words (nānārthavarga), all other synsets in a class show some semantic relation, to the class it belongs to and sometimes even to the preceding or following synsets.
- These semantic relations indicate various kinds of relations.
- They may be classified as hierarchical or associative.
- Hierarchical relations
  - hypernym - hyponym
  - holonym - meronym
- Associative relations
  - association among human beings
  - association of certain objects with certain other objects.

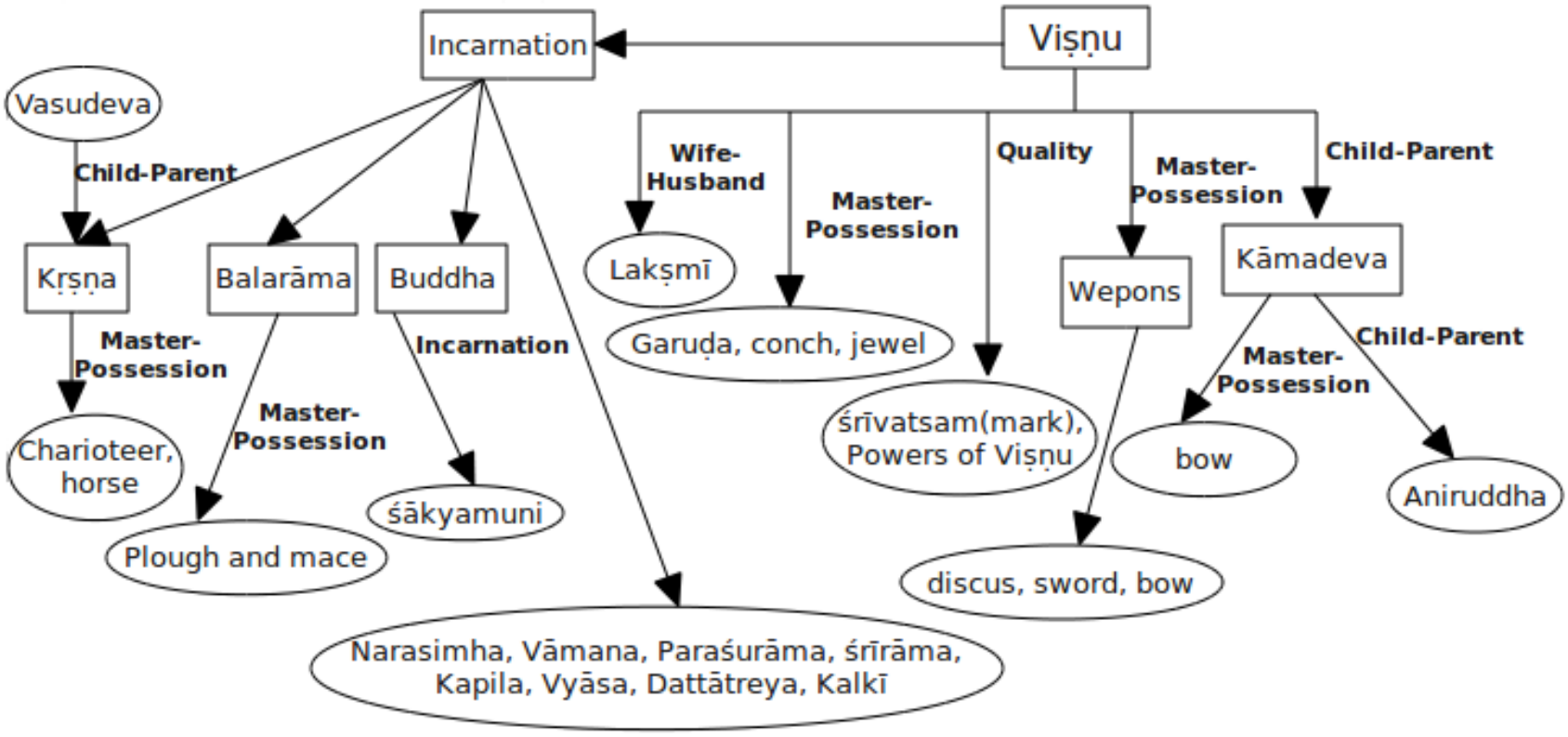
Knowledge structure in Sanskrit Kośas

- Except for the polysemous words (nānārthavarga), all other synsets in a class show some semantic relation, to the class it belongs to and sometimes even to the preceding or following synsets.
- These semantic relations indicate various kinds of relations.
- They may be classified as hierarchical or associative.
- Hierarchical relations
  - hypernym - hyponym
  - holonym - meronym
- Associative relations
  - association among human beings
  - association of certain objects with certain other objects.

Example 1(a): Viṣṇuḥ from Amarakośa (1.1.18 - 1.1.29)



Example 1(b): Viṣṇuḥ from Vaijayantīkośa



Example 2(a): Samayaḥ from Amarakośa (1.4.1 to 1.4.9)

Time (1.4.1)
- Lunar day (1.4.1)
  - First lunar day (1.4.1)
- {Day (1.4.2)
  - Morning (1.4.2 - 1.4.3)
  - Twilight (1.4.3)
  - Evening (1.4.3)
  - First four hours of a day (1.4.3)
  - Second four hours of a day (1.4.3)
  - Third four hours of a day (1.4.3)
  - Period of the day (1.4.3)
  - Night (1.4.3 - 1.4.4)
    - A dark night (1.4.5)
    - A moonlight night (1.4.5)

cont....

A night and two days (1.4.5)
First part of night (1.4.6)
Midnight (1.4.6)
Sequence of nights (1.4.6)
Space of three hours (1.4.6) }
Last day of the half month (1.4.7)
Precise moment of the full or the new moon (1.4.7)
Full moon day (1.4.7)
Full moon whole day(1.4.8)
Full Moon with a little gibbous on part of a day (1.4.8)
No moon day (1.4.8)
wanning crescent (1.4.9)
No moon whole day (1.4.9)

Example 2(b): Samayaḥ from Vaijayantīkośa (2.1.79 - 2.1.81)

half of a lunar month (*pakṣa*) (2.1.79)
first (white) *pakṣa* of a lunar month (2.1.79)
second (black) *pakṣa* of a lunar month (2.1.79)
pakṣa in which the *tithis* become shorter (2.1.79)
pakṣa in which the *tithis* become longer (2.1.79)
short fifteenth day of the second *(kṛṣṇa)pakṣa* (2.1.80)
short first day of a *pakṣa* (2.1.80)
lunar month (2.1.80)
solar month (month in which the sun passes to another *rāśi*) (2.1.80)
star month (2.1.81)
month of 30 days (2.1.81)

Example 3: Kṣatriyaḥ from Amarakośa (2.8.1 to 2.8.10)

- Man of the military tribe (2.8.1)
  - King (2.8.1)
    - An emperor (2.8.2)
    - Universal monarch (2.8.2)
    - King over a country (2.8.2)
    - Paramount sovereign (2.8.3)
  - Multitude of kings (2.8.3)
  - Multitude of military tribe (2.8.4)
    - Minister (2.8.4)
      - Deputy minister (2.8.4)
    - Priest (2.8.5)
    - Judge (2.8.5)
    - King's companions (2.8.5)
    - Body guards of a king (2.8.6)

cont....

Warder (2.8.6)
Superintendent (2.8.6)
Village Superintendent (2.8.7)
Superintendent of many villages (2.8.7)
Superintendent of Gold (2.8.7)
Superintendent of Silver (2.8.7)
Superintendent of the womens' appartments (2.8.8)
Outside guard of the womens' appartment (2.8.8)
Attendant of a king (2.8.9)
Eunuch (2.8.9)
Prince whose territories lie on the frontiers of those of the enemy (2.8.9)
Neighboring prince (2.8.9)
Prince whose territories lie beyond those of the friend (2.8.10)
Enemy in the rear (2.8.10)

Implicit relations

These were three samples from three distinct topics involving totally different kind of relations. All these are semantic in nature. A more detailed study of such examples showed that following relations occur more frequently.

- *avayavāvayavī-sambandha* (part-whole relation)
- *parāparājāti-sambandha* (is a kind of relation)
- *janyajanaka-bhāva* (child-parent relation)
- *patipatnī-bhāva* (husband-wife relation)
- *svasvāmi-bhāva* (master-possession relation)
- *ājīvikā* (livelihood)

There are a few other relations such as ādhāra-ādheya, vaṃśa-vaṃśīya, etc. which occur less frequently.

Relational statistics

The occurrence of various relations in terms of Head-Words and all the words belonging to the synsets denoted by these head words is shown in Table.

No.	Relation	Headwords	Words
1	is_a_kind_of	2239	6807
2	is_a_part_of	560	1654
3	janya-janaka	17	193
4	sva-svāmī	36	122
5	ājīvikā	30	106
6	pati-patnī	25	105

WordNet and Amarakośa

- A pilot study to compare the Sanskrit synsets as defined in Amarakośa with the Hindi synsets from Hindi WordNet was carried out.
- The HindiWordNet had 27,879 synsets and Amarakośa has 4017 sysets.
- Out of these, only 1554 concepts 'matched' at the conceptual level.
- This matching was done 'manually'.
- Over a period of time, the word meanings undergo changes due to various factors.
- According to the speaker's or listener's intension the meaning of a word may change.

WordNet and Amarakośa

- A pilot study to compare the Sanskrit synsets as defined in Amarakośa with the Hindi synsets from Hindi WordNet was carried out.
- The HindiWordNet had 27,879 synsets and Amarakośa has 4017 sysets.
- Out of these, only 1554 concepts 'matched' at the conceptual level.
- This matching was done 'manually'.
- Over a period of time, the word meanings undergo changes due to various factors.
- According to the speaker's or listener's intension the meaning of a word may change.

WordNet and Amarakośa

- A pilot study to compare the Sanskrit synsets as defined in Amarakośa with the Hindi synsets from Hindi WordNet was carried out.
- The HindiWordNet had 27,879 synsets and Amarakośa has 4017 sysets.
- Out of these, only 1554 concepts 'matched' at the conceptual level.
- This matching was done 'manually'.
- Over a period of time, the word meanings undergo changes due to various factors.
- According to the speaker's or listener's intension the meaning of a word may change.

WordNet and Amarakośa

- A pilot study to compare the Sanskrit synsets as defined in Amarakośa with the Hindi synsets from Hindi WordNet was carried out.
- The HindiWordNet had 27,879 synsets and Amarakośa has 4017 sysets.
- Out of these, only 1554 concepts 'matched' at the conceptual level.
- This matching was done 'manually'.
- Over a period of time, the word meanings undergo changes due to various factors.
- According to the speaker's or listener's intension the meaning of a word may change.

WordNet and Amarakośa

- A pilot study to compare the Sanskrit synsets as defined in Amarakośa with the Hindi synsets from Hindi WordNet was carried out.
- The HindiWordNet had 27,879 synsets and Amarakośa has 4017 sysets.
- Out of these, only 1554 concepts 'matched' at the conceptual level.
- This matching was done 'manually'.
- Over a period of time, the word meanings undergo changes due to various factors.
- According to the speaker's or listener's intension the meaning of a word may change.

WordNet and Amarakośa

- A pilot study to compare the Sanskrit synsets as defined in Amarakośa with the Hindi synsets from Hindi WordNet was carried out.
- The HindiWordNet had 27,879 synsets and Amarakośa has 4017 sysets.
- Out of these, only 1554 concepts 'matched' at the conceptual level.
- This matching was done 'manually'.
- Over a period of time, the word meanings undergo changes due to various factors.
- According to the speaker's or listener's intension the meaning of a word may change.

- The meaning-evolution may be classified into three broad categories.
  - saṁkoca(shrinking)
  - vistāra(expansion)
  - ādeśa(imposition)
- Though Hindi is descended from Sanskrit, and bears a lot of commonality with original Sanskrit words with regards to the meanings, yet we see many instances where it has undergone various kinds of changes such as expansion, shrinking, etc. This is the main cause behind the un-matched synsets.
- Nevertheless this study was useful to link the Hindi synsets with those of Amarakośa, and it should be extended further to other Indian Languages as well.
- Such a study will also give a fair idea about how various languages have deviated from Sanskrit.

- The meaning-evolution may be classified into three broad categories.
  - saṁkoca(shrinking)
  - vistāra(expansion)
  - ādeśa(imposition)
- Though Hindi is descended from Sanskrit, and bears a lot of commonality with original Sanskrit words with regards to the meanings, yet we see many instances where it has undergone various kinds of changes such as expansion, shrinking, etc. This is the main cause behind the un-matched synsets.
- Nevertheless this study was useful to link the Hindi synsets with those of Amarakośa, and it should be extended further to other Indian Languages as well.
- Such a study will also give a fair idea about how various languages have deviated from Sanskrit.

- The meaning-evolution may be classified into three broad categories.
  - saṁkoca(shrinking)
  - vistāra(expansion)
  - ādeśa(imposition)
- Though Hindi is descended from Sanskrit, and bears a lot of commonality with original Sanskrit words with regards to the meanings, yet we see many instances where it has undergone various kinds of changes such as expansion, shrinking, etc. This is the main cause behind the un-matched synsets.
- Nevertheless this study was useful to link the Hindi synsets with those of Amarakośa, and it should be extended further to other Indian Languages as well.
- Such a study will also give a fair idea about how various languages have deviated from Sanskrit.

- The meaning-evolution may be classified into three broad categories.
  - saṁkoca(shrinking)
  - vistāra(expansion)
  - ādeśa(imposition)
- Though Hindi is descended from Sanskrit, and bears a lot of commonality with original Sanskrit words with regards to the meanings, yet we see many instances where it has undergone various kinds of changes such as expansion, shrinking, etc. This is the main cause behind the un-matched synsets.
- Nevertheless this study was useful to link the Hindi synsets with those of Amarakośa, and it should be extended further to other Indian Languages as well.
- Such a study will also give a fair idea about how various languages have deviated from Sanskrit.

- The meaning-evolution may be classified into three broad categories.
  - saṁkoca(shrinking)
  - vistāra(expansion)
  - ādeśa(imposition)
- Though Hindi is descended from Sanskrit, and bears a lot of commonality with original Sanskrit words with regards to the meanings, yet we see many instances where it has undergone various kinds of changes such as expansion, shrinking, etc. This is the main cause behind the un-matched synsets.
- Nevertheless this study was useful to link the Hindi synsets with those of Amarakośa, and it should be extended further to other Indian Languages as well.
- Such a study will also give a fair idea about how various languages have deviated from Sanskrit.

- The meaning-evolution may be classified into three broad categories.
  - saṁkoca(shrinking)
  - vistāra(expansion)
  - ādeśa(imposition)
- Though Hindi is descended from Sanskrit, and bears a lot of commonality with original Sanskrit words with regards to the meanings, yet we see many instances where it has undergone various kinds of changes such as expansion, shrinking, etc. This is the main cause behind the un-matched synsets.
- Nevertheless this study was useful to link the Hindi synsets with those of Amarakośa, and it should be extended further to other Indian Languages as well.
- Such a study will also give a fair idea about how various languages have deviated from Sanskrit.

- The meaning-evolution may be classified into three broad categories.
  - saṁkoca(shrinking)
  - vistāra(expansion)
  - ādeśa(imposition)
- Though Hindi is descended from Sanskrit, and bears a lot of commonality with original Sanskrit words with regards to the meanings, yet we see many instances where it has undergone various kinds of changes such as expansion, shrinking, etc. This is the main cause behind the un-matched synsets.
- Nevertheless this study was useful to link the Hindi synsets with those of Amarakośa, and it should be extended further to other Indian Languages as well.
- Such a study will also give a fair idea about how various languages have deviated from Sanskrit.

ConceptNet and the Sanskrit Kośas

- There is a fundamental difference between the development of ConceptNet and the knowledge found in the Sanskrit Kośas.
- The ConceptNet is aimed at 'capturing the common sense'.
- This common sense typically concerns with the behaviourial observations, social norms etc.
- It focuses typically on actions with which other concept nodes are related.
- The Sanskrit kośas on the other hand contain words which are:-
  1. culture specific (svarga, viṣṇu, etc.)
  2. reveal social or man made structures (kingdom, house, etc.)
  3. throw light on various social practices

ConceptNet and the Sanskrit Kośas

- There is a fundamental difference between the development of ConceptNet and the knowledge found in the Sanskrit Kośas.
- The ConceptNet is aimed at 'capturing the common sense'.
- This common sense typically concerns with the behaviourial observations, social norms etc.
- It focuses typically on actions with which other concept nodes are related.
- The Sanskrit kośas on the other hand contain words which are:-
  1. culture specific (svarga, viṣṇu, etc.)
  2. reveal social or man made structures (kingdom, house, etc.)
  3. throw light on various social practices

ConceptNet and the Sanskrit Kośas

- There is a fundamental difference between the development of ConceptNet and the knowledge found in the Sanskrit Kośas.
- The ConceptNet is aimed at 'capturing the common sense'.
- This common sense typically concerns with the behaviourial observations, social norms etc.
- It focuses typically on actions with which other concept nodes are related.
- The Sanskrit kośas on the other hand contain words which are:-
  1. culture specific (svarga, viṣṇu, etc.)
  2. reveal social or man made structures (kingdom, house, etc.)
  3. throw light on various social practices

ConceptNet and the Sanskrit Kośas

- There is a fundamental difference between the development of ConceptNet and the knowledge found in the Sanskrit Kośas.
- The ConceptNet is aimed at 'capturing the common sense'.
- This common sense typically concerns with the behaviourial observations, social norms etc.
- It focuses typically on actions with which other concept nodes are related.
- The Sanskrit kośas on the other hand contain words which are:-
  1. culture specific (svarga, viṣṇu, etc.)
  2. reveal social or man made structures (kingdom, house, etc.)
  3. throw light on various social practices

ConceptNet and the Sanskrit Kośas

- There is a fundamental difference between the development of ConceptNet and the knowledge found in the Sanskrit Kośas.
- The ConceptNet is aimed at 'capturing the common sense'.
- This common sense typically concerns with the behaviourial observations, social norms etc.
- It focuses typically on actions with which other concept nodes are related.
- The Sanskrit kośas on the other hand contain words which are:-
  1. culture specific (svarga, viṣṇu, etc.)
  2. reveal social or man made structures (kingdom, house, etc.)
  3. throw light on various social practices

ConceptNet and the Sanskrit Kośas

- There is a fundamental difference between the development of ConceptNet and the knowledge found in the Sanskrit Kośas.
- The ConceptNet is aimed at 'capturing the common sense'.
- This common sense typically concerns with the behaviourial observations, social norms etc.
- It focuses typically on actions with which other concept nodes are related.
- The Sanskrit kośas on the other hand contain words which are:-
  1. culture specific (svarga, viṣṇu, etc.)
  2. reveal social or man made structures (kingdom, house, etc.)
  3. throw light on various social practices

ConceptNet and the Sanskrit Kośas

- There is a fundamental difference between the development of ConceptNet and the knowledge found in the Sanskrit Kośas.
- The ConceptNet is aimed at 'capturing the common sense'.
- This common sense typically concerns with the behaviourial observations, social norms etc.
- It focuses typically on actions with which other concept nodes are related.
- The Sanskrit kośas on the other hand contain words which are:-
  1. culture specific (svarga, viṣṇu, etc.)
  2. reveal social or man made structures (kingdom, house, etc.)
  3. throw light on various social practices

ConceptNet and the Sanskrit Kośas

- There is a fundamental difference between the development of ConceptNet and the knowledge found in the Sanskrit Kośas.
- The ConceptNet is aimed at 'capturing the common sense'.
- This common sense typically concerns with the behaviourial observations, social norms etc.
- It focuses typically on actions with which other concept nodes are related.
- The Sanskrit kośas on the other hand contain words which are:-
  1. culture specific (svarga, viṣṇu, etc.)
  2. reveal social or man made structures (kingdom, house, etc.)
  3. throw light on various social practices

- In addition to the observational facts such as classification of animals, the flora and fauna etc.
- This structure thus involves many non-observational facts, which are mostly part of our culture.
- In this sense the structure of these kośas is complementary to what the ConceptNet provides and is an essential part of the Indian culture.
- There is a necessity to make these valuable resources available in suitable e-form so that the NLP community working in Indian Languages can be benefited.

- In addition to the observational facts such as classification of animals, the flora and fauna etc.
- This structure thus involves many non-observational facts, which are mostly part of our culture.
- In this sense the structure of these kośas is complementary to what the ConceptNet provides and is an essential part of the Indian culture.
- There is a necessity to make these valuable resources available in suitable e-form so that the NLP community working in Indian Languages can be benefited.

- In addition to the observational facts such as classification of animals, the flora and fauna etc.
- This structure thus involves many non-observational facts, which are mostly part of our culture.
- In this sense the structure of these kośas is complementary to what the ConceptNet provides and is an essential part of the Indian culture.
- There is a necessity to make these valuable resources available in suitable e-form so that the NLP community working in Indian Languages can be benefited.

- In addition to the observational facts such as classification of animals, the flora and fauna etc.
- This structure thus involves many non-observational facts, which are mostly part of our culture.
- In this sense the structure of these kośas is complementary to what the ConceptNet provides and is an essential part of the Indian culture.
- There is a necessity to make these valuable resources available in suitable e-form so that the NLP community working in Indian Languages can be benefited.

Amarakośa-jñāna-jāla : Proof of concept

- The Amarakośa-jñāna-jāla is developed as a web application.
- The application provides a search result of a query.
- This search result is generated dynamically from -
  - The structured lexicon of the Amarakośa
  - The supplementary tables marking the relations.

Amarakośa-jñāna-jāla : Proof of concept

- The Amarakośa-jñāna-jāla is developed as a web application.
- The application provides a search result of a query.
- This search result is generated dynamically from -
  - The structured lexicon of the Amarakośa
  - The supplementary tables marking the relations.

Amarakośa-jñāna-jāla : Proof of concept

- The Amarakośa-jñāna-jāla is developed as a web application.
- The application provides a search result of a query.
- This search result is generated dynamically from -
  - The structured lexicon of the Amarakośa
  - The supplementary tables marking the relations.

Amarakośa-jñāna-jāla : Proof of concept

- The Amarakośa-jñāna-jāla is developed as a web application.
- The application provides a search result of a query.
- This search result is generated dynamically from -
  - The structured lexicon of the Amarakośa
  - The supplementary tables marking the relations.

Amarakośa-jñāna-jāla : Proof of concept

- The Amarakośa-jñāna-jāla is developed as a web application.
- The application provides a search result of a query.
- This search result is generated dynamically from -
  - The structured lexicon of the Amarakośa
  - The supplementary tables marking the relations.

Structured Lexicon

- The main structured lexicon consists of synsets stored in the form of a set of records.
- Each record corresponds to a word in the Amarakośa (excluding the meta-language words).
- It consists of 5 fields as
  - Stem
  - Amarakośa index
  - Gender
  - Vargah
  - Head Word

Structured Lexicon

- The main structured lexicon consists of synsets stored in the form of a set of records.
- Each record corresponds to a word in the Amarakośa (excluding the meta-language words).
- It consists of 5 fields as
  - Stem
  - Amarakośa index
  - Gender
  - Vargah
  - Head Word

Structured Lexicon

- The main structured lexicon consists of synsets stored in the form of a set of records.
- Each record corresponds to a word in the Amarakośa (excluding the meta-language words).
- It consists of 5 fields as
  - Stem
  - Amarakośa index
  - Gender
  - Vargaḥ
  - Head Word

Structured Lexicon

- The main structured lexicon consists of synsets stored in the form of a set of records.
- Each record corresponds to a word in the Amarakośa (excluding the meta-language words).
- It consists of 5 fields as
  - Stem
  - Amarakośa index
  - Gender
  - Vargaḥ
  - Head Word

Structured Lexicon

- The main structured lexicon consists of synsets stored in the form of a set of records.
- Each record corresponds to a word in the Amarakośa (excluding the meta-language words).
- It consists of 5 fields as
  - Stem
  - Amarakośa index
  - Gender
  - Vargaḥ
  - Head Word

Structured Lexicon

- The main structured lexicon consists of synsets stored in the form of a set of records.
- Each record corresponds to a word in the Amarakośa (excluding the meta-language words).
- It consists of 5 fields as
  - Stem
  - Amarakośa index
  - Gender
  - Vargaḥ
  - Head Word

Structured Lexicon

- The main structured lexicon consists of synsets stored in the form of a set of records.
- Each record corresponds to a word in the Amarakośa (excluding the meta-language words).
- It consists of 5 fields as
  - Stem
  - Amarakośa index
  - Gender
  - Vargaḥ
  - Head Word

Structured Lexicon

- The main structured lexicon consists of synsets stored in the form of a set of records.
- Each record corresponds to a word in the Amarakośa (excluding the meta-language words).
- It consists of 5 fields as
  - Stem
  - Amarakośa index
  - Gender
  - Vargaḥ
  - Head Word

Stem

- Amarakośa lists words in nominative cases.
- Nominal stems are the contents of this field.
- In case of feminine words, this field contains the feminine stem, (stem after adding the feminine suffix).
- In case of nānārthavarga, the polysemous word is entered in this field.
- The reason for choosing nominal stem over the nominative case form is the ease in linking the Amarakośa words with the existing computational resources such as morphological analysers and generators and various e-lexicons, which typically expect a prātipadikam and not a prathamānta (ending in nominative case).

Amarakośa index

- This field contains a reference to an entry in the Amarakośa as a 5 tuple of numbers, separated by dots.
- The 5 numbers in the 5 tuple refer to the kāṇḍa, varga, śloka, pāda and the word number respectively.

Liṅgam (gender)

- This field contains the gender of the stem.

Vargaḥ

- This field contains the name of the varga of the token.

Head Word

- The first four fields cover all the explicit information that can be easily extracted automatically.
- The marking of synonymous words is obvious only through the world knowledge or through the commentaries.
- To provide a handle to each synset, we created a field termed as *Head Word* which provides a name to each synset.
- Head Words are unique and act as a reference ID for a synset.
- The total number of Head Words give us the total number of synsets in the Amarakośa.
- We denote the synset corresponding to a Head Word *W* by *Syn(W)*.
- The choice of Head-Words is mainly guided by the Bhānuji Dīkṣitā's *Sudhā* commentary.
- When a better choice was available in the Malayalam commentary *Triveṇī* or *Pārameśvarī* or the Hindi commentary *Prabhā*, it was chosen.

Example of the Data entry

This is an example of a śloka 2.5.5 converted to a structured table.

Token	Reference	Gender	Varga	Head-Word
śivā	2.5.5.1.1	strī	siṃhādivargaḥ	jaṃbhūkaḥ
būrimāya	2.5.5.1.2	puṃ	siṃhādivargaḥ	jaṃbhūkaḥ
gomāyu	2.5.5.1.3	puṃ	siṃhādivargaḥ	jaṃbhūkaḥ
mṛgadūrtaka	2.5.5.1.4	puṃ	siṃhādivargaḥ	jaṃbhūkaḥ
śṛgāla	2.5.5.2.1	puṃ	siṃhādivargaḥ	jaṃbhūkaḥ
vañjaka	2.5.5.2.2	puṃ	siṃhādivargaḥ	jaṃbhūkaḥ
kṛoṣṭu	2.5.5.2.3	puṃ	siṃhādivargaḥ	jaṃbhūkaḥ
pheru	2.5.5.2.4	puṃ	siṃhādivargaḥ	jaṃbhūkaḥ
pherava	2.5.5.2.5	puṃ	siṃhādivargaḥ	jaṃbhūkaḥ
jaṃbuka	2.5.5.2.6	puṃ	siṃhādivargaḥ	jaṃbhūkaḥ

Tables marking various relations

- The relations are among various Head Words and are marked as records.
- Each record corresponds to one synset ID.
- The first field of each record consists of the synset ID, and remaining six fields correspond to the Head Words that bear a relation of those six relations with the synset ID in the first field.
  - avayavāvayavi-sambandha (is_a_part_of)
  - parāparājāti-sambandha (is_a_kind_of)
  - janya-janaka-bhāva
  - pati-patnī-bhāva
  - sva-svāmi-bhāva
  - ājīvikā

Examples of relations

Relation R	**Head-Word W**	**Related Head-Word W'**
is_a_part	rātrimadhyaḥ	rātriḥ
is_a_kind	gaṅgā	nadī
Janya-janaka	indraḥ	jayantaḥ
Pati-patnī	lakṣmī	viṣṇuḥ
Sva-svāmi	garuḍaḥ	viṣṇuḥ
Ājīvikā	dhīvaraḥ	matsyaḥ

Implementation

- From the structured lexicon table and the table of relations we build data bases using the built-in dbm engines of unix.
- The programes are written in Perl.
- These dbm engines use hashing techniques to enable fast retrieval of the data by key.
- Three hash tables are built from the structured lexicon.
  - Head-word hash
    Key=stem and Value=head-word
  - Synset hash
    Key=head-word and Value=synset
  - Word-info hash
    Key=stem and Value=word-index and gender
- From the table of relations, corresponding to each relation R, we built a hash table which returns a head-Word W associated with another head-word W', if W' is related to W by relation R.

Web application

- *Amarakośa-jñāna-jāla* is presented as a web application.
- It is developed with **apache** web server and **perl** for CGI script.
- The query is in the form of - a word and a relation.
- Machine produces all the words related to the given word by the chosen relation.
- The word here may be either a stem or an inflected word form.
- In the case of inflected word form, machine consults the morphological analyser to get the stem.
- When a cursor is placed on a word a tool tip shows its word-index and gender.

Example of a synset

पर्यायवाची(Synset)

शृगाल

अर्थः :: जम्भूकः | वर्गः :: सिंहादिवर्गः | भूरिमाय, गोमायु, जम्बुक, क्रोष्टु, मृगधूर्तक, फेरव, फेरु, शिवा, शृगाल, वञ्चक, शालावृक

काण्ड,वर्ग,श्लोक,पाद :: 2.5.5.1.3,लिङ्ग :: पुं.

Conclusion

This implementation may serve as a model to build similar tools for various other kośas like -

Nāmamālā, *Śabdaratnākara*, *Śabdacandrikā*, *Nānārthasaṅgraha*, *Anekārthadhvanimaṅjarī*, *Viśvaprakāśa*, *Abhidhānaratnamālā*.

The Amarakośa is now available with various kinds of search facilities as a web service at

http://sanskrit.uohyd.ernet.in/~anusaaraka/sanskrit/samsaadhanii/amarakosha/home.html

Dhanyavādāḥ