GENERAL INFORMATICS & SANSKRIT

BA SANSKRIT

V SEMESTER

CORE COURSE

(2011 Admission)

UNIVERSITY OF CALICUT

SCHOOL OF DISTANCE EDUCATION

CALICUT UNIVERSITY P.O. MALAPPURAM, KERALA, INDIA - 673 635

UNIVERSITY OF CALICUT
SCHOOL OF DISTANCE EDUCATION

STUDY MATERIAL

BA SANSKRIT

(2011 Admission Onwards)

V SEMESTER

CORE COURSE

GENERAL INFORMATICS & SANSKRIT

PREPARED BY: DR. E. JAYAN
Associate Professor
Department of Sa nskrit
VTB College, Mannampetta,
Palakkad

SCRUITNISED BY: DR. K. JAYANARAYANAN
Associate Professor
Department of Sa nskrit
Sree Kerala Varma College,
Thrissur.

***LAYOUT & SETTINGS*:** COMPUTER CELL, SDE

CONTENT

MODULE- I

COMPUTER BASICS

1. EARNING GOALS

In this chapter we will learn:

1. The concept of an algorithm
2. The major parts of a computer required to carry out algorithms.
3. The main characteristics of computers.
4. How problems are solved by computers

Computers are now affecting every sphere of human activity and bringing about many changes in industry, government, education, medicine, scientific research, law, social sciences and even in arts like music and painting. They are presently used, among other applications, to

- Design buildings, bridges and machines
- Control space vehicles
- Assist in railway reservation
- Control inventories to minimize material cost
- Grade examinations and process results
- Aid in teaching
- Systematically store and quickly retireve data on land records
- Play games like chess and video games

The areas of application of computers are confined only by limitations on human creativity and imagination. In fact any task that can be carried out systematically, using a precise step-by-step method, can be performed by a computer. Therefore it is essential for every educated person today to know about a computer, is strengths, it s weaknesses and its internal structure. The main objective of this book is to explain these in a simple language.

1.1. ALGORITHMS

In order to solve a problem using a computer it is necessary to evolve a detailed and precise step-by-step method of solution. Step-by-step methods for solving problems are not new or peculiar to computer. They have been in use for a very long time, and in almost all walks of life. One such method, taken from a popular magazine, is given as Example 1.1.

Example 1.1 Recipe for potato bondas

Ingredients

Potatoes 250gms, Chopped onion 3 (small), Finely chopped chillies 5, Gram flour 100gms, oil for frying, water for batter, Salt 2 teaspoons.

Method

Step 1 : Boil potatoes till cooked, peel and mash them until they are soft.

Step 2 : Mix fried onions, green chillies and salt with the mashed potatoes

Step 3 : Take little portions of the mixture and make small balls.

Step 4 : Mix gram flour, water and a little sale and beat well till a smooth and creamy batter is obtained.

Step 5 : Dip the potato balls in the batter. Take out and deep fry in oil on low fire.

Step 6 : Take out when the balls fried to a golden brown colour.

Result

A dozen potato ready to be served hot with tomato sauce.

The receipe given above has the following properties:

1. It begins with a list of ingredients which we may call the inputs.
2. A sequence of instructions is given to process the inputs.
3. As a result of carrying out the instructions, some output (namely, Potato bondas) are obtained.

The instructions given to process the inputs are, however, not precise. They are ambiguous. For example, in step 4, we are instructed to " beat well till a smooth and creamy batter is obtained". The interpretation of "smooth and creamy" can vary from person to person. Due to such imprecise instructions, different persons following the same receipe with the same inputs can produce potato bondas which differ in size, shape and taste!

We will now examine another step-by-step procedure (again taken from a popular magazine).

Example 1.2 A procedure to knit a sweater

Materials required

Needless No.12-2, wool 4 ply – 9 balls.

Method

Step 1 : Cast on 133 stitches

Step 2 : Repeat Steps 3 and 4, 11 times

Step 3 : Knit 2, * Purl 1, Knit 1, Repeat from * to last stitch, Knit I.

Step 4 : Knit 1, * Purl 1, Knit 1, Repeat from * to End.

……

……

(similar Steps)

Result

A sweater.

The above example illustrates the following points :

1. This procedure has inputs, a set of steps to process the inputs to produce an output.
2. The procedure is more precise and unambiguous than the receipe for potato bondas.

   There is very little chance for misinterpretation.
3. Different types of instructions used in the procedure are very few. If one knows how to knit, how to purl, cast stitches on or off needless, and count, then one can knit any sweater.
4. By a proper permutation and combination of this elementary set of instructions a virtually infinite number of patterns for sweaters may be created. For example, if Step 3 is made "knit 1, *Knit 1, purl 2, Repeat from * to last stitch" the pattern of the sweater will be entirely different.

The preciseness of the instructions combined with their small variety makes it possible to design a machine which can knit automatically. In fact, a forerunner of the modern computer was a loom designed by a French engineer, called Jacquard, in 1801, which could be 'programmed' to create a large number of patterns. The program consisted of cards with specific patterns of holes in them which controlled the loom.

We have illustrated two simple step-by-step methods of solving problems. We will now explain when a step-by-step method can be called an algorithm.

The origin of the word algorithm is not known. It is , however, generally accepted among mathematician that it comes from the name of a famous Arab mathematician Abu Jafar Mohammed ibn Musa al- Khowarizmi (circa 825) (literally meaning father of Jafar Mohammed, son of Moses, native of al – Khowarizm) who wrote the celebrated book " Kitab at jabr Walmuqabla" (Rules of Restoration and Reduction). The last part of his name al-Khozwarizmi was corrupted to algorithm. An algorithm may be defined as a finite sequence of instructions (to solve a problem) which has the following five basic characteristics:

1. An algorithm begins with instruction(s) to accept inputs. These inputs are processed by subsequent instructions in the algorithm.
2. The processing rules specified in the algorithm must be precise and unambiguous. In other word, the instructions must not be vague. It must also be possible to carry them out. For example, the instruction "Go to hell" is precise, but cannot be carried out.

3. Each instruction must be sufficiently basic such that it can, in principle, be carried out by a person with paper and pencil.
4. The total time to carry all the steps in the algorithm must be finite. As algorithms may contain instructions to repetitively carry out a group of instructions this requirement implies that the number of repetitions must be finite.
5. An algorithm must produce one or more outputs (namely the result of processing the inputs).

Based on the above definition we see that the recipe of Example. 1.1. does not qualify as an algorithm as it is not precise. The knitting pattern, on the other hand, does qualify.

We will now evolve a step-by-step procedure to solve a data processing job. The problem is to find the average number of vowels occurring in a short passage.

Example 1.3 Procedure to find the average number of vowels in a passage

Step 1 : Let number of Characters = 0

Step 2 : Let number of vowels =0

Step 3 : Repeat Steps 4,5,6 and 7 until end of passage is reached.

Step 4 : Road one character from the passage.

Step 5 : Add 1 to number of Character.

Step 6 : If the character is any one of the letters A,E,I,O,U, a,e,i,o,u, add I to number of vowels.

Step 7 : Move to next character.

Remark : step 8 is reached when no more characters are left in the passage.

Otherwise we go back to step 4.

Step 8 : Average Number of Vowels $= \frac{\text{Number of vowels}}{\text{Number of characters}}$

Step 9 : Write Average Number of vowels, Number of characters.

Step10 : Stop.

This step-by-step procedure qualifies as an algorithm as :

(i) It has an input, namely, the passage to be examined.
(ii) The processing steps are precisely specified.
(iii) Each instruction is basic and can be carried out by a person with paper and pencil.
(iv) The procedure terminates when the end of passage is reached. The total number of steps carried out is finite.

1.2. A SIMPLE MODEL OF A COMPUTER

If a machine is to be built to carry out the algorithm of Example 1,3 it should have the following features :

(i) It should be able to interpret each instruction in the algorithm and carry it out.

(ii) It should be able to read a text consisting of a set of characters.

(iii) It should be able to add and divide.

(iv) It should be able to compare a character read from the passage with the letters A,E,I,O,U, a,e,i,o,u and if it matches any one these characters, to add 1 to the number of vowels.

(v) It should be able to store (or remember) computed values.

(vi) It should be able to write (or output) the answers obtained.

(vii) If the steps in the algorithm are to be carried out automatically without human intervention the entire algorithm must be stored for reference. As the algorithm is also repetitive, that is, a set of steps are carried out again and again until a condition is satisfied , it is necessary to have all the steps in the algorithm stored in a storage unit.

A computing machine designed to carry out algorithms for data processing thus has the configuration of Fig.1.1. Referring to Fig.1.1. it is seen that an input unit is provided to read the algorithm and the data to be processed by the algorithm. The memory unit stores.

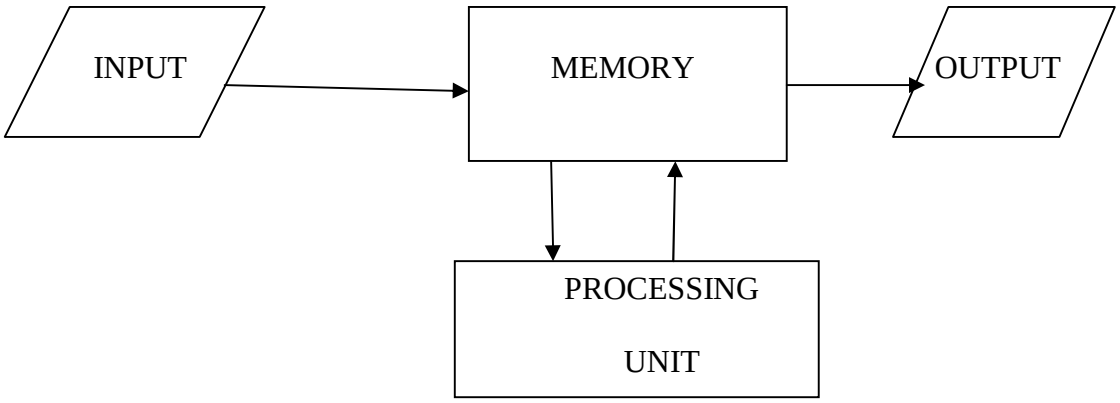

Fig.1.1 Organization of a computer.

The algorithm and computed values. The processing unit interprets the instructions and carries them out. It has the capability to perform arithmetic operations, character manipulation operations, and logical operations. The output unit prints or displays computed results.

In Example 1.2, it was seen that it is possible to develop an enormous number of different interesting patterns by permuting and combining a few basic

types of instructions. The same principle is used in building computers. Thus by using a computer's processing unit which can interpret and execute as few as ten different operations, it is possible to perform a large variety of data processing tasks.

1.3. CHARACTERISTICS OF COMPUTERS

The interesting features of a computer are :

1. Computers are built to carry out a small variety of instructions. It is not necessary to have more than about 100 distinct instructions even for a very powerful machine.
2. Instructions are extremely simple; e.g., add, subtract, read a character, write a character, compare numbers, characters, etc.
3. Most instructions are carried out in less than a millionth of a second.
4. Instructions are carried out obediently with no questions asked.
5. Instructions are carried out without any mistakes.

A computer may thus be thought of as a servant who would carry out instructions, precisely, obediently, uncritically, at a very high speed, and without exhibiting any emotions. As human beings, we use judgments based on experience, often on subjective and emotional considerations. Such value judgments' often depend on what is called sound "commonsense". As opposed to this, a computer exhibits no emotions and has no commonsense. An algorithm may be written for a computer to compose music based on rules of composition, but the cannot judge the quality of the resultant music. It must be clearly understood that computers are machines which can be programmed to follow instructions; they don't have their own priorities and judgments. Computers are machines which can help mankind in many ways; but they do not threaten us.

Being obedient without exercising 'commonsense' can be very annoying and unproductive. This is illustrated by the experience of a Colonel who sent his obedient peon to a post office with the order "Go to the Post Office and buy ten 25 paise stamps but not ordered to return with them!

A consequence of the uncritical acceptance of orders by a computer is the need to give extensive, detailed, and correct instructions for solving problems. This can be quite challenging.

MODULE- II

THE INTERNET HAS ARRIVED

The World is Changing

A revolution is taking place. It started quietly and has grown to involve much of the world. On an average day, the following events may occur:

- A young couple use a computer to plan a road trip from their home in Ohio to a friend's house in San Francisco, California. Within seconds, they receive a description of the shortest route along with maps and detailed descriptions of distances and earth turn.
- An investment broker in Houston, Texas sits down at a personal computer and runs a program that accesses current prices on the New York Stock Exchange. After looking at the list, the broker purchases shares of two stock and sells shares of another.
- A class of elementary school children in Chicago, Illinois use a computer network to read a newspaper article from the New York Times.
- A mother living in Maryland uses a computer to view the weather in North Carolina where her daughter lives. After seeing that the storm has passed, she sends a note expressing her relief.
- A teenager in Seattle, Washington uses a computer to listen to music from new CD. Later, he uses the computer to place an order.
- A grandparent in Boston, Massachusetts uses a computer to inquire about airline flights, make a reservation, and purchase a ticket for a trip to visit a grand child.
- A group of company executives hold a meeting. One executive is in New York. Another in Florida, and a third is vacation I Colorado. Each sits in front of a computer that has both a camera and microphone attached. They see pictures of one another on the screen and hear each others' voices.
- A computer program runs at 6.00 PM in Atlanta, Georgia to send a copy of a company's daily sales receipts to a branch office in London, England.
- A high school student in Taiwan uses a computer to see and hear a tour of the campus at a university in Hawaii, including a live video showing students walking across campus with palm trees swaying in the breeze. Later, the students use the computer to send a fax to a relative who is visiting Australia.

What do all these events have in common? In each, people are using the Internet, a communication system that is revolutionizing the way we work and play.

If you have not yet encountered the Internet, you will soon. Let's look at some statistics:

- The Internet currently reaches hundreds of millions of people in over, 209 countries (every populated area on the planet).
- Most of 2-year and 4-year U.S. colleges and universities have access to the Internet; many public schools also have access.
- The U.S. military has been using Internet technology for over twelve years; it played a role in military actions as early as Operation Desert Storm in the early 1990s.
- Scientists have been using the Internet since 1980.
- The U.S. President and the White House are accessible via the Internet as are other government agencies in many countries.

Numbers Do Not Tell the Story

The most common assessment of the Internet's significance measures the number of computer that connect to it. However, conventional computer connections tell only part of the story. The Internet reaches ships at sea, planes in the air, and mobile vehicles on land. Private companies provide access to Internet services through the telephone and able television systems, making it possible to reach the Internet from any home or office that has a telephone or cable TV.

To assess the impact of the Internet, one might as, " what has it affected?" The answer is, "Almost everything."

So the question becomes:

The Internet has arrived ; are you ready for it?

Learning About the Internet

This book answers the question "what is the Internet?" in the broadest sense. It examined the origins of computer networking and its application to everyday problems. It focuses on the services that the Internet provides and helps the reader understand their importance. More important, it removes some of the mystery and helps the reader understand how the technology works.

Learning about the Internet is not something one can complete in an afternoon – learning never stops because the Internet keeps changing. The Internet is similar to a newsstand – when new information appears, it replaces older information. Each time you visit the newsstand or the Internet, you can find something new.

Of course, information on the Internet changes much more rapidly than information in a conventional newsstand. In fact, because information on the

Internet comes from computers and automated systems, it can change instantly. For example , if one accesses weather information twice in a single minute, the information obtained from the two accesses can differ because computers can measure weather and change he report constantly.

In addition to resembling a newsstand, the Internet also resembles a library because it has tools that aid the search for information. In a traditional library, for example, one finds a card catalog and a reference desk. The Internet has similar services that help one find information electronically.

Understanding the Big Picture

Grasping all the details of the Internet is impossible because the Internet continues to change. Thus, no one can know the locations of all the interesting data or the way to obtain the lowest price for an item. More important, because new applications are being invented, no one can obtain a complete description of all the services available. Finally, because individual computers and software programs differ, one cannot expect the same details to apply to all computers.

To avoid becoming overwhelmed with details, we will examine the fundamental of the Internet. Instead of focusing on how use a particular computer, a particular brand of software, or a particular Internet service, we will consider the basics of how the Internet works and how information services use the basic mechanisms. In essence, we will examine the capabilities of the Internet'.

Understanding Internet capabilities makes it much easier to read computer manuals and to use the Internet. In particular, because most computer manuals specify the details of how to accomplish a task without describing why one needs to perform the task beginners often find them difficult to follow. Knowing how the Internet works and the purpose of each service helps put the details in perspective.

Terminology and Technology

A complex technology, the Internet has spawned a terminology that can be daunting. This book clearly explains the Internet technology using analogies and examples. It shows how the pieces fit together, emphasizing basics instead of details. It discusses the services that the Internet offers, explains the flexibility they provide, and describes how they can be used.

More important, this book introduces technical terminology used for computer networking and the Internet. Instead of providing a list of terms, early chapters present definitions in a historical perspective that shows how communication systems evolved. For example, early chapters explain the difference between digital and analog information. Instead of using computer networks as an example, the chapters relate the terminology to everyday experience.

Growth And Adaptability

Part of the mystique surrounding the Internet arises from its rapid success. While the Internet has grown, dozens of other attempts to provide the

same services have failed to deliver. Meanwhile the Internet continues to expand by adapting to change, both technical and political. We will examine why Internet technology has worked so well and how it has adapted to accommodate change.

Another amazing part of the Internet story is its incredible growth. We will look at how the Internet continues to grow and the consequence of such growth.

The Impact of the Internet

Perhaps the most significant aspect of the Internet is its impact on society. Once restricted to a few scientists, it is quickly becoming universal. It reaches governments, businesses, schools, and homes worldwide. We will examine how the Internet changes people's lives, and what we can expect in the future. In summary, the rest of this book looks at what the Internet is and what it can do for you.

Organization of The Book

This book is organized into four sections. After a brief tour of Web sites in chapter 2, the first section (Chapter 3 through 7) introduces communication system concepts and terminology. If you already understand digital and analog communication, universal service, and binary data encoding, you may choose to skim this section. The second section (chapter 8 through 11) reviews the history of the Internet and its incredible growth. The third section (chapters 12 though 19) describes basic Internet technology and capabilities. It examines how Internet hardware is organized and how software provides communication. Be sure to understand this section ; it provides the foundation for later chapters. The final section describes services currently available on the Internet. For each service, it explain both how the service works and how it can be used.

A Personal Note

I still remember an occasion several years ago when a colleague bluntly asked me the question, "what is the Internet?" I had been involved with Internet research for many years, and had written a popular college textbook that described the Internet and the principles underlying its design. I knew many details about the hardware and software systems that comprised the Internet, how the computers were connected, and the details of communication. I also knew most of the researchers who were working on technical improvements. What puzzled me most was that the person asking already knew basic technical details and had a copy of my textbook. What more could I say?

As I contemplated the question, my colleague guessed that I misunderstood and said, "I do not want to know about computers and wires. I mean, in a larger sense, what is Internet, and what is it becoming? Have you notices that it is changing? Who will be using it in ten years? What will they do with it?

The questions were importance because they pointed out a significant shift. Early in its history, most of the users of Internet were the experts who helped build it. The Internet has outgrown its research beginnings and has become a

powerful tool. It is a facility used by almost everyone. It is being used in ways that the experts had not imagined.

Getting Started: Hands – On Experience

Introduction

The material in this text gives a broad, conceptual view of Internet capabilities. It explains what the Internet can do and how the underlying technology operates, without focusing on details. For example, the text does not attempt to provide a catalog of the most interesting items available on the Internet, nor does it discuss how to use any particular computer or particular brand of software. Thus, the entire text can be understood without direct access to a computer.

Despite the emphasis on concepts, readers who have Internet access are encouraged to use the Internet as they read. Hands – on experience provides intuition and familiarity that enhances learning in the same way that access to a car helps one better appreciate learning about highways and destination reachable on those highways. To aid readers in getting started, this short chapter suggests some possibilities. It describes the concept of a Web browser, which is a computer program, and provides a few examples of sites to visit. Chapter 22 revisits the topic of browsers and explains in greater detail how they work. Readers already familiar with the Internet need not spend time reviewing; they can skip directly to the next chapter.

The web : Sites And Pages

Although many applications have been devised that use the Internet, one particular application has evolved that incorporates the best features of others through a single interface. The application is known as the World wide Web or merely “The web”.

Most large organizations and many smaller ones have a Web Site. We think of an organization’s Web Site as its point of contact – the site contains all the information about the organization that is available to Internet users. For example, the Web site for a corporation might contain a catalog of products the company offers, price, instructions for ordering, a list of the company’s employees along with their telephone numbers and a statement of their responsibilities, or information about employment possibilities. In addition to corporations, institutions, such as schools and governments also maintain Web Sites. Finally, individuals maintain Web sites that contain personal information.

To prevent a user from being overwhelmed by a large volume of information, the contents of a Web site are divided into “pages”, In general, an Internet user views one page of information at a time. Consequently, a page is usually designed to fit on a user’s screen. When a user moves to a new page, the contents of the previous page are no longer visible; to view the previous page again, the user must explicitly return to it.

Web Browsers and Browsing

Access to the World Wide Web requires a computer, a connection to the Internet, and special software on the computer. In later chapters, we will learn that two types of software are required : Internet communication software and application software. The communication software speaks the "language" of the Internet, making it possible for the computer to communicate with other computers. The application software handles all interaction of with the user. When the user makes a request for information, the application software uses the Internet to access the requested information, and then displays the result on the screen for the user to see.

An application program used to access the World Wide Web is known as a Web browser, and someone who uses such a program is said to be browsing the web. Although several browser programs exist, two completely dominate the market : Netscape's Navigator and Microsoft's Internet Explorer. The two differ only in minor details. Interestingly, both are free.

Using a Browser

The paradigm for browsing the Web is straightforward. To begin a user connects his or her computer to the Internet and invokes a browser. When the browser runs, it creates a window on the screen. The major portion of the window consists of an area that the browser uses to display a Web page (until a user specifies otherwise, a browser is likely to display an advertisement). A small area near the top of the browser's window contains controls that a user invokes to specify a Web page. Within the control area, the browser has a small text area labeled location that is used to specify a particular Web Site. Each site is given an identification string called a URL (later chapters explain the terminology ; for now, the details are unimportant). To visit a Web Site, the user moves the mouse to the location box, enters the URL for the desires site, and presses the RETURN key. The browser contacts the specified site and displays site, the main page of information. The page can contain a mixture to text and graphic images.

Once a Web page appears on the screen, moving to another page is easy. Some of the items on a page are "links" to other pages. Text that corresponds to a link is usually displayed in a different color and is underlined; images that correspond to a link do not always have a clear indication. To follows a link, one uses the mouse to move the cursor over the item on the screen and clicks. The browser automatically uses the Internet to obtain the page that corresponds to the link, and then replaces the display with the new page. Thus browsing consists of entering the URL for a specific site and then following links from one page to another.

Examples of Web Sites and Services

A few examples will help demonstrate the type of information available on the Web. The selection is not meant to imply that these are preferred sites; it merely gives an idea of the scope and variety of information available. To visit any of the sites, enter the URL that is given.

An Online Newspaper

The New York Times, one of the most well – known newspapers in the world, maintains a Web site that contains current newspaper articles and stock market information. The URL is :

www.nytimes.com/

Weather Information

The Weather Channel has an online service that gives weather information and forecasts. The URL is :

www.weather.com/

Driving Directions

Several Web sites offer detailed driving directions that specify routes, distances, and turns needed to drive from one address to another. For example, the Mapquest service can be bound at:

www.mapquest.com/

Satelite Maps

An Internet site exists that has satellite and aerial maps of the earth's surface. You should be able to see a satellite map of your neighbourhood. The URL is :

www.terraserver.com/-

Radio Stations

Many ratio stations send audio over the Internet. Stations, categorized by music format can be found by following the links on the page given by URL :

www.broadcast.com.radio/

An Online Fashion Mail

Every large city has a fashion mall with stores that carry well-known brands. An online version can be found at :

www.fashionmall.com/

A Retail Bookstore

Amazon is one of the best known online retailers. Initially an online bookstore, it now offers other merchandise. The URL is :

www.amazon.com/

Stock Information

The New York Stock Exchange site provides a description of the investment community as well as information about popular stock index values and stock prices. The URL is :

www.nyse.com/

Music Clips

Chapter 26 explains how audio is delivered over the Internet, and notes that MP3 is a popular format used to encode music. A selection of music and other audio encoded in MP3 format can be found at:

www.mp3.com/

Electronic Postage

It is possible to purchase electronic postage and use it to send a conventional letter. In essence, electronic postage operates like a postage meter instead of a conventional stamp – after paying for postage, the customer prints a label that the post office honors in place of stamp. The URL is :

www. E-stamp.com/

Satire

Many web sites contain satirical material. A satire of Microsoft arranged to took like a board game can be found at :

www.ms- monopoly.com/

Summary

This chapter briefly reviewed the most popular Internet service, the World Wide Web. We learned that information found at given Web Site is divided into pages ; a user views one at a time with a computer program known as a browser Each page can contain text and pictures, or can connect to audio and video. We also learned that a wide variety of information is accessible through the Web. Chapters 22 through 25 will examine Web technologies in more detail and explain how the Internet Provides such services.

Before The Internet
A gentle Introduction To
Communication Systems
Concepts And Terminology

Telephones Everywhere

Introduction

This Chapter introduces the concept of universal service. It uses a familiar example to show how the assumption of universal service can affect our view of a communication service, and explains why the Internet is becoming a necessity as it becomes universal.

A Communication Service

The Internet is communication technology. Like the telephone before it, the Internet makes it possible for people to communicate in new ways ; However to the average person living now, digital communication is as novel as telephone communication was to the average person living one hundred years ago. We can learnt many lesson from the story of telephone service that apply directly to the Internet.

Selling Communication

To understand how a new communication technology infiltrates society, think back approximately a century. Imagine yourself as a salesperson in an average town in the U.S. who has the job of selling telephone service.

All things considered, the economic times you face are full of promise. Excitement and optimism pervade industry. After all, society is experiencing an industrial revolution. Everywhere you find that mechanization has replaced manual labor. The steam engine has replaced water wheels and animals as a source of power, some industries are starting to use engines that run on gasoline. Factories are producing more goods than ever before.

Of course, a telephone salesperson of a century ago would have had little or who firsthand experience using a telephone. Indeed, he or she may have had only a few demonstrations before going out to sell telephone service.

Imagine that you walk into a small company and talk to the owner about telephone service. What can you say? You could tell the owner that the company needs a telephone because it will allow customers to place orders easily. Or you could say that a telephone will allow employees to check with suppliers, order raw materials, or trace shipments that do not arrive on schedule. May be you would ask the owner if he or she goes out to lunch with other business owners, and point out that a luncheon could be arranged in a few seconds over a telephone. You could say that a telephone is easy to use. Or, you might take a more serious approach and point out if fire struck the business, a telephone could be used to reach the firehouse; it might save property or lives.

How do the owners react to your telephone sales pitch? Some are interested ; many are skeptical. A few are delighted, but others are angry. Although some will think the idea has merit, many will laugh. Some want to redesign business practices, but most resist. A few want a telephone just because it is new and lends status to their establishment. Despite what they say, most owners believe that they will continue to conduct business without using a telephone.

Limited Access

Selling Telephone service without having used it can be difficult. But let's make the task of selling easier. Suppose that you had grown up in a world with telephone service, and that you had used telephones all your life. Then suppose that you were transported back in time almost 100 years and tried to sell telephone service. You might think that it would be easy to convince people to buy telephone service knowing how it can be used, but you would be surprised by what you face.

The first shock you encounter when trying to sell telephone service is learning that the service of a century ago did not work the same way as modern telephone service. Back then, telephone service meant local service. Each town or village decided independently when to run wires, hire a switchboard operator, and establish phone service. More important, each town chose a telephone systems exited, they were incompatible – running wires from one town to the next did not guarantee that the telephone systems in the two towns could would together. From a business perspective, even if a company installed a telephone, it could not be used to order supplies from other parts of the country. You quickly discover.

Having an independent local telephone service in each town limits the usefulness of a telephone.

High Cost

The second shock you encounter when trying to sell telephone service approximately a century ago is learning that even when it is available, telephone services is expensive. An average family cannot afford a telephone in their home. In addition to buying the telephone itself, many telephone companies charge each subscriber the true in a given neighborhood than to enlist additional subscribers. More important, for a large part of the population who live in rural areas, telephone service is out of the question.

After many attempts to sell telephone service to individuals fail, you report back to your employer with a conclusion.

Telephone service will not be a viable business until the cost of service becomes low enough for an average family to have a phone installed.

The Difficult Transition

In a without telephones, convincing a business to install one may seem impossible. If the business cannot use it to call suppliers in remote parts of the country and local customers do not have easy access to a telephone, there is little economic justification for acquiring one. In fact, after thinking about the world of

telephone service that we enjoy and the world of telephone service approximately a century ago, you realize ;

The single most important idea behind a communication service arises from its coverage – if no one else has the service it is useless ; if everyone else has the service, it is a

necessity.

The transition between the two extremes is difficult. it requires businesses and individuals to invest in a new communication technology before the economic benefit is obvious. If they choose a technology that does not catch on they lost their investment. Even of others adopt the technology, it may have insufficient subscribers to justify economically. Many people remain reticent when a new technology arrives. They wait to see what everyone else will do hoping to minimize their financial risk. The financial decision is more difficult for a business, which must decide how rapidly to install telephones. If the business has too few phones, callers will receive a busy signal; if the business has too many, the phones sit idle, meaning that the business has wasted resources.

Ubiquitous Access

Why did everyone in the U.S. eventually choose to subscribe to telephone service? If you are a student of history, you know the answer : because the U.S. government decided that upiquitous telephone service was important for the country. The governments of most other countries reached the same decision. The U.S. government established a regulated monopoly. American Telephone and Telegraph (AT & T) . It mandated that telephone services be providing a single, large system.

Because one company owned and operated much of U.S. telephone network, many tasks were easy. For example, AT & T could specify the technical details to how the phone system in one city interconnected with the phone system in another. Having one company own the system made it easy to deploy new technology. A single company also made it easy to define a global numbering system so that a subscriber in one city could directly dial the telephone number of a subscriber in another city.

In short, the result of the government action was universal telephone service available at price an average family could afford. Within a few decades, most businesses and a large portion of the population could be reached by telephone. Of course,universal telephone service could have occurred without government intervention ; we can only speculate about what might have happened. The important point is no that the government intervened , but that popularity of the telephone surged as universal service became business procedures. As business and individuals started acquiring telephones, it became apparent to everyone that telephones were important. Acquiring one became a necessity. Telephone service changed from a luxury reserved for the rich to something expected by the average family.

In the U.S., the telephone system became the communication system of choice in the twentieth century because the government mandate of universal telephone service guaranteed that subscribing would benefit everyone.

Relevance to The Internet

Like the telephone system, the Internet provides communication. Currently, the Internet falls in the awkward transition period between limited access and universal service. Although connections are growing rapidly, the Internet does not reach everyone. Although the U.S. government has contributed to Internet development, it has not decided to mandate universal service. Thus, unlike the phone system, Internet growth has relied on economics. As a result growth has proceeded in a haphazed manner. In the mid-1990s, major business decided that they would benefit from an Internet connection and began to mention their web pages in general advertising. By 2000, millions of homes had internet access, and businesses were upgrading both their computers and Internet connections to handle the increased traffic.

During the transition convincing someone who has not used the Internet that it offers exciting new possibilities is like trying to sell telephone service before a universal phone system is in place. Often people who see Internet technology smile politely and nod, while thinking to themselves. “That’s all very nice, but how would I use it?” The question is the analogy of someone who has never seen a telephone asking. “Yes, I see how it works, but whom would I call?”

The answer of course is that once everyone is connected to the Internet, you will want to use it to contact businesses, friends, banks, schools, government offices, and relatives. A later chapter discusses, Internet growth, and shows that the Internet is becoming universal.

MODULE- III

RELEVANCE OF SASTRAS FOR NATURAL LANGUAGE PROCESSING

P. RAMANUJAN

0. ABSTRACT

In this paper, the issues in Natural Language Processing (NLP) which could possibly benefit from sastraic studies are discussed. Only the basic are touched upon and a few practical implementation issues are detailed. The DESIKA package developed at centre for Development of Advanced Computing (C-DAC), Poona is referred to throughout to clarify/support discussion.

Word, sentence and discourse (Syntactic, Semantic and Pragmatic) levels are considered. Knowledge base consisting of data and rules are covered. Issues regarding Generation and Analysis modes of plain and accented Sanskrit input processing are briefly explained. Subanta, tinanta, krdanta and taddhita forms etc. are discussed. Implementation of simple sentence analysis is described.

Semantic analysis includes generic and specific syntactic semantic mappings, activity and conceptual classification, ontological characterizations, sentence, coherence factors etc. drawing upon from the Sastras, Computer outputs from DESIKA package are available.

1.0. INTRODUCTION TO SANSKRIT AND NLP

Ever since Comparative Philology and Indology became subjects of serious and significant studies (for over a century now) Linguists have paid attention to Paninian grammar in ample measure and Navya Nyaya theories of meaning, sentential import etc. to a good extent. With the advent of computational methods of analysis, computational Linguistic has become an area of advanced research.

In Artificial Intelligence (AI), issued like Processing of natural (Conversational) Languages, Representing knowledge in computers, particularly inference and modeling commonsense/ contextual knowledge are the frontier areas of research. Here ambiguity is 'the' issue while interpreting (and, of course, during generation also). Added to this is the factor vivaksa (speaker's intention') for the determination of which hardly any help is available. Thus, this is essentially multi – disciplinary involving cognitive sciences as well, since language comprehension requires human competence at lexical, syntactic, semantic, phonetic, prosodic, cognitive and socio-contextual levels. As an example for a computer to 'understand' a simple sentence like " the master teaches scriptures to ascetic disciples" all the factors aforesaid are to be available in the machine.

In Indian Sastras (scientific treatises), these factors are dealt with integrally and the result is the formulation of a comprehensive system of language description for correct usage. This system has Vedas at the nucleus and a host of auxiliary sciences to assist their understanding and preservation. These are really devices for ensuring distortion – free transmission over generations, through oral tradition. This system has remarkably served the cause for millennia, and that is precisely why there is interest in modern scientists working in AI about these literary traditions (details later).

Currently the three major sastras, viz. Nyaya Vyakarana and Mimamsa are studied with great interest in search of clues regarding the above mentioned issues. Some of the recent efforts in this direction are : Paninis grammar (auto – semantics) as cognitive knowledge structure in 'Simurg' project in Germany, two – level morpho – phonology of Sanskrit in Finland, Morphological analysis of Sanskrit by computer in Netherlands, Paninian database project in U.S. and National Language Understanding (NLU) systems based on Sabdabodha theories described in the Sastras in India. The C- DAC, Pune, has developed GIST (Graphics and Intelligence based script Technology) card which provides facility for processing linguistic data in any of the Indian scripts (and also many foreign scripts). DESIKA, a package developed for study of Sanskrit works in GIST environment (details later).

2.0. ISSUES IN NLP

Many of the problems conforming NLP boil down to three of the central issues in AI, i.e., representation, reasoning and recognition. It is essential to devise means to tackle these issues as AI seeks to give a computational account of intelligent behavior, which should include understanding communication. Also, much of world's knowledge is set down in natural language and if it is to be used by artificial manipulation.

3.0. CONTENTS AND RELEVANE OF SASTRAS

Traditionally knowledge is classified topically into subjects called Vidyasthanas. These include four Vedas (scriptures), their six limbs (Vedangas) and four supplementaries (Upangas). While Vyakarana is a limb, Nyaya and Mimamsa are supplementary to Vedas. Vedas are the treasure – house of all knowledge and are proverbially infinite and eternal. They are chiefly classified into Rk, Yajus, sama and Atharva Vedas. Understanding them thoroughly is not possible without the six limbs which deal with the factors as shown below in the order (name – meaning – linguistic factor dealth – limp type, explanation).

Nirukta, etymology and exegesis (lexical) – Ears, provides, a repertoire of words (accented) or philogocal explanation of difficult Vedic words; Vyakarana grammar (syntactic) – mouth, helps generate innumerable grammatically valid words – forms from a finite set of roots through rules. Siksa science of pronunciation (phonetic) – nose, classifieds sound phonetically and defines pronunciations and euphonic combination including accents for the character set of the language ; Chandas metrics /prosody (prosodic)- feet, defines intonation structure for speech to communicate emotions appropriately ; jyaustia astronomy

(temporal and spatial) – eyes, define the suitable occasions for proper results ; kalpa, ceremonial practical – hands , prescribes rituals and rules for ceremonial and sacrificial acts.

Among the four supplementary sciences, Nyada deals with semantics by a system of logical compatibility and validity rules ; Mimamsa gives guidelines for interpreting Vedic texts and deals with discourses and philosophy, puranas details the mythological and spiritual aspects of life while Dharmasastras lay down moral codes of rectitude.

Of these, the themes which are found to be directly relevant and readily usable as yet, are the structure of grammar of Panini as a rulebase for natural language generation (and syntactic analysis) and the method of sabdabodha (semantic extraction) described in Vyakarana (at word level), Nyaya (at sentence level) and Mimamsa (at discourse level). However, when vedic analysis is to be done, which is of paramount importance to clearly understand the Sastraic system, the others are also inevitable.

3.1. The Vyakarana Sastra

This science with rules by Panini contains well structured description of Sanskrit grammar and deals with word – level and sentence level systactic aspects. Both plain and accented forms are exhaustively treated. Besides the rules called Astadhayayi, there are database for ganapatha (of nominal stems), dhatupatha (of verbal roots), lingansasana (gender determination) and siksa (phonetics). The theory of karaka (functional relationship) relating the case structure of syntax and the denoted objects is a major outcome of universal application to natural languages. The significant aspect of the activity denoted by the verbal root : dhatvartha mukhya – visesyaka – sabdabothah.

3.2. The Nyaya Sastra

This science with rules by sage Gautama deals with padartha Vibhaga (conceptual classification of things) as pramana and prameya (means and objects of knowledge). Theories of validity and error are also covered. A model structure of technical languages for unambiguous description (Navya – Nyaya) is a by product. The significant aspect of the import of a sentence according to scholars of this Sastra is normally the word in nominative case (agent) of the action denoted by the sentence: prathamantartha mukhya – visesyaka – sabdabothap.

3.3. The Mimiamsa Sastra

This science with rules by sage Jaimini establishes principles for understanding the contents of Vedic passages by maxims and pragmatic reasoning. Contextual factors and common sense aspects are formulated with hierarchy of knowledge sources for conflict resolution and disambiguation. Various means of practical wisdom are a result of these formulations. Moral codes and jurisprudence are derived from some of these in Indian Culture. The significant aspect of the import of a sentence according to scholar of this sastra is

normally the bhavana (intention or force typically in injunction or commands). Bhavana mukhya viseyaka – sabdabodhah.

4.0. COMPUTATIONAL KNOWLEDGE BASE

To be able to utilize knowledge contained in natural language in free form within a machine, certain formats are to be devised and the knowledge is to be represented suitably in a machine understandable form. It is to be noted that the machine can only process 0's and I's and any information is to be made available in this form ultimately. To achieve this conversion of free form in natural languages to format machine form, many techniques are available. The Sastras can also assist in devising novel means for the purpose. Thus any price of information pertaining to knowledge of any type, has to be stored in proper form and this stored and formatted information is termed knowledge base. This is usually divided into two parts. Le., the database (information processed dynamically or variable) and the rulesbase (static information which is applied on the database). So, Kosa (lexicon) and Sutra. (rules) traditionally (Astadhayyi jaganmata amarakoso jagatpita) correspond to the two parts of knowledge base. (A knowledge base usually exists in alphanumeric form. Of character strings and numeric codes.

4..1. SYNTATIC (GRAMMATICA) DATABASE

As said earlier, the dynamically processed (for variable) information could consists of the fundamental lexical units of languages which are processed for generation or analysis. Thus, prakrti (stem or root) and prataya (affix prefix in – fix of suffix) can be the basic data units. Lexicon like Amarakosa can be thought of as the database for nominal stems and indeclinable and dhatupatha for verbal roots. We also treat the various affixes as data and stores them accordingly (Pratyaya Kosa).

4.11. Prakrti Kosa (BASE LEXICON)

This lexicon would have information fields of pratipadika (nominal stem), ending, gender (s) declension type (explained later), derivation type, meaning, concept type etc. for all the words of say, Amarakosa typically. Of these fields of information the word alone need be in character form. While therest of the details could be in a suitably coded form. Similarly, for verbal roots, the dhatupatha could be had in fields of roots with it as a character field and information regarding the gana, dhatu and it svars, padi etc. in a coded form. Avyayas (indeclinables are also 'coded' about their type like upasarga, gatinipata, karmapravachaniya, kidanta, taddhitanta etc.) formation, base and suffix if any meaning category etc. One way of 'coding' attempted in DESIKA would be discussed in implementation.

4.111. Amarakosa (VOCABULARY)

As an example of base lexicon, we could have a database of the Amarakosa in its original form (i9.e., verses) and have programs to interpret and process the data or we could also have a hybrid scheme with both coded and original forms and provide quotation from source with reference, as needed. Here, the

technical definitions (Samjnas) of the author can be interpreted by program functions, e.g., rupacheda, sahacarya visesa vidhi etc. by condition checks, trisu, dvayoh, sesartha, tvanta, athadi ect. By interpreting functions. WE can also analyse the text using thee meta – rules or definitions. Other Kosas (lexicons) can also be similarly handled.

4.112. Ganapatha (LIST OF NOMINAL STEM CLASSES)

Among the nominal base lexicons, Paninis ganapatha lists various stems with a certain syntactic property, particularly, regarding tattlhita (secondary derivatives) etc. These happen to be the data when we program taddhita generation or analysis. Here, membership of a particular list and its characteristics are of interest and this information can be easily coded.

4.113 Dhatupatha (LIST OF VERBAL ROOTS)

Paninis dhatupatha is the source of verbal root data. We can code information regarding gana, padi, it karmitya, idagama, dhatu svara and it svara etc. explicity or derive them using Astadhyayi sutras by program functions. Other dhatupathas can also be in character form while the rest of the details can be in numeric code form. The effect of upasargas (preverbs) on root meanings would also have to be stored as data.

4.12 Prataya Kosa (AFFIX LEXICON)

This includes nominal declensional, verbal conjugational and modal, primary and second\ary derivational affixes, feminie suffixes etc. in a categorized manner. Here also, the affixes are character fields while their characteristic parameters could be coded. Semantics are also included in terms of relationships, properties etc. For example, 'sup' (nominal declensional) suffixes denote case relationships and number, 'tin' (verbal conjugational and modal) affixes denote temporal, modal, active, personal and numeric relationship of relate with the verbal root meanings. Krt and taddhita suffixes (primary derivatives) denote karaka (functor) sesa (residual or generic) and patronymic relationships etc. as specified in karaka, vibhakti and taddhita prakaranas (sections) in Astadhyayi.

4.13. OTHER (SEMANTIC) DATABASES

Besides grammatical data, other aspects of linguistic data mentioned earlier are also to be suitably available in the machine. Some of there are detailed in the following sections. The main point here is that the sastraic system in exhaustive and comprehensive in nature such that by appropriately representing its multi – dimensional character, we can build a good aid to its thorough study thorough the advanced computational processes. It will also help define a knowledge representation system based on ancient Indian tradition and comparatively study the modern development in its light.

4.131. Padartha Vibhaga (CONCEPTUAL DATA)

The conceptual categories and ontology established in Nyaya Ssastra could be stored as semantic or logical database, in the form of frames etc. The base lexicon can be integral with the semantic features as well as the manual

dictionaries contain) or there can be a separate semantic lexicon, in which case, the linking of the syntactic and semantic lexicons has to be provided.

4.132. Sastrantara Kosa (APPLICATION DOMAIN DATA)

At discourse level, for different applications, contextural and domain specific information about the concerned field would be needed. Thus, a database of the particular branch of study is to be established with links to syntactic and semantic lexicons. Mimamsa sastra happens to be very important from the point of view of explaining the concepts in contextural and pragmatic reasoning. The formulation of its database is thus essential. Similarly, preparation of database for other vidyasthana also could be undertaken.

4.133. Vaidika Kosa (VEDIC DATABASE)

For the purposes of Vedic processing, certain additional databases pertaining to accented words and text would be needed, thus the Nighantu of Yaska (which is a database of assorted Vedic words) the Nirukta (which explains the Nighantu etymologically) in terms of derivative processes, the parameters of pronunciation and phonetics given in siksa works for the character set of the language, the sakha (Vedic branch) specific special characters etc, from pratisakhya works and the entire vedic text in padapatha form as the 'fundamental corpus' are required to be established.

4.134 Sahitya Kosa (LITERARY DATABASE)

For the study of classical Sanskrit treatises like Rasagangadhara, Kavyaprakasa or works like Sisupalavadha, Anargharghava, Raghyvamsa the respective texts will have to be keyed in and a database created . Here through a lexical update program one can only update those entries that are not already part of the lexicon, dynamically (even while analyzing i.e. on line.

4.135 SECONDARY DATABASE

Certain additional databases in the form of technical terms, definitions, tables of correspondence and mappings, query words and their expected answer types, statistics, links between various databases, schemes of alphanumeric coding for the different parameters of the database etc. e.g., all sastriya samjuna (paribhasika sabdas) subanda paradigm types of so on are also to be prepared. These are mainly guided by programming considerations.

4.2 RULESBASE

The rule base is a set of linguistic instructions for processing the input data. This could be in the form of if (condition) them action rules, or a simple or complex network linking input and output states or certain concepts or descriptions. Most of syntactic processed may be of rules form, the semantic relationships being graphically or conceptually defined. Only empirical principles may be all that is possible in pragmatic levels of processing.

4.21 GRAMMATICAL RULES BASE

Grammar of Sanskrit deals with syntactic details are word and sentence levelws, mainly for the spoken form. There is one to one correspondence between

spoken and written forms. Svaras (accents – different emphases) of sentence and euphonic combination of words are also covered. These are available from astadhyayai arranged topically and process wise. The processes are basically adesa (substitution), agama (augmentation) or lopa (elision) at lexeme, phoneme or morpheme levels.

4.211 Astadhayai

This is the well known descriptive rule base of Sanskrit grammar by sage Panini. There are eight adhyayas (chapters) with four padas (quarters) each and having a total of about 4000 suras (aphorisms) The rules are extremely brief, require meta rules to interpret and fill epllipses and are very closely knit in nested chaining form). The rules are of samjna (definitive) paribhasa (interpretative) adhikara and anuvritti (jurisdicational) vidhi (operative) apavada (exceptional), niyama (restrictive) or atidesa (extensive) types. These rules can be effected by suitable functions in program and the process carried out. The trace of the process of generation of recognition is also possible to be given by the computation process.

4.212 Linganusasma

The rules pertaining to determination of gender of words are listed in Paninis Linganusasana which has simple rule – exception – extention sequence and contains 191 rules in all and lists about 1200 nominal stems. Amarakosa also has these factors covered in lingadi – sangraha varga. With these rules programmed, the user can just specify a pratipadika (nominal stem) and get the valid declension through the system itself, including multiple genders.

4.22. OTHER RULESBASES

The semantic cognitive and socio – contextural aspects are not easily amenable for simple formation in the machine. Here, the concerned Sastras also have evolved and adopted appropriate techniques which are very relevant to A1 and NLP. These include classification of things of characteristics parameters logical compatibility criteria truth and error value definition and philosophy, sentential coherence issues, propriety conditions etc. detailed thoroughly in Nyaya and Mimamsa Sastras. These are also to be stored suitably in the machine for semantic and contextual analysis, conflict resolution and disambiguation. Common sense formulation world knowledge representation etc. need these details badly.

5.0. PROCESSING STAGES / LEVELS

We now detail certain stages of computational processing of linguistic data pertaining to case – based languages like Sanskrit. We consider both generation and analysis (recognition) at word, sentence and discourse levels. Parsing a given sentence or creating a sentence to denote indicated information is the test of machine understanding of natural language. This understanding may demonstrated by query processing, voice change or paraphrasing the input sentence for a parser. Conversely parser output should be input to the generator to obtain the input sentence.

5.1. WORD LEVEL

Generating inflectional natural language words and recognizing finished words in terms of base and affix (including multiple possibilities) require computational knowledge base aforesaid and a proper algorithm (step by step method) for arriving at the output from the input, systematically. This procedure would then be capable of capturing the traditional expertise in its exhaustive form for all future uses and refinements. Thus, 'knowledge' can be preserved in contra-distinction to 'data' or 'information' storage and retrieval.

5.11 Nispattih (GENERATION)

For generating natural languages words, we need as the input for base, a nominal stem or verbal root and for affix, a specified semantic feature like functional (case) or personal relationship etc, for derivative forms, the significant criteria are to be specified. Here, the machine offers invaluable help in listing all possible choices that can be indicated by using suitable database and structures of formats. All the rules involved in the process of generation together with a trace of all intermediate results on application of each rule or condition is obtainable. This feature would be invaluable in academic efforts like study, research etc. We now consider various word types in Sanskrit.

5.111 Subanta (NOMINAL)

Nominal declension take archetypal terminating suffixes denoted by the siglum 'sup' and are inflected into eight possible cases and three numbers. Thus, the nominal stem and the desired case and number are to be input for generation. Here, ending character (vowel or consonant) of the stem and its gender are the two factors that influence the way the archetypal 'sup' suffixes get modified. Other parameters like pronouns, certain derivative types and specific syntactic criteria also caused special modifications to the suffixes. Thus, the declension is according to the applicable paradigm type; the process of internal combination of stem and suffix also may entail certain modification (e.g. cerebralisation) in inflected form. It is thus possible to establish all distinct declensional paradigm types before and by consulting the rulebase and use them in generation. We have identified 327 distinct paradigm types. Works like Sabda Retnavalli could be generated by the machine automatically at will with explanations.

5.112. Tinanta (VERBAL)

Verbal declension depend on root parameters like the conjungation (or group), indicatory characters, accents, mode, voice, tense or mood, person and number. They may also be affected by preverbs if any. Here the archeptyal affixes are denoted by the siglum 'tin' The input is a root with indicatory characters alongwith the desired choice of preverbs mode, voice, tense/mood,person and number. Here also, generalization of the modification of the archeptyal affixes into paradigm types may be attempted theoretically. However, our studies have shown not much savings in efforts on this count as the diversity is high (1213 paradigm types for around 2000 roots). It is thus best

generated, ab – initio using rules. Works like Dhaturupakosa could be generated by the machine automatically at will, explanations.

5.113. Krdantas (PRIMARY DERIVATIVES)

For Krdantas, the choice of verbal root, preverbs, if any and the desired krt suffixes (with anubandhas) are input. The effects of these are applied through rulebase and the necessary processing is carried out (including sandhis). A paradigm type analysis of krt suffixes on semantic basis would help represent meanings for further stages. Thus, the meaning could also be indicated instead of the desired suffix (action sequence, certain functor, relationships or material properties etc.) These are highly expressive words forms covering participal infinitive substantive, adjective gerundial types etc. Even works like *krdantarupamala* can be generated automatically when the system is completely developed with the added advantage of providing 'trace' the rules involved and intermediate forms.

5.114 Taddhitas (SECONDARY DERIVATIVES)

For Taddhitas choice of nominal stem (from lexicon, ganapatha etc.) and desired taddhita suffix are input. Here also, grouping of taddhita suffixes on functional basis can be carried out and this information is used to select the desired form. List of taddhita forms are thus generated.

5.`115. Samasa (COMPOUNDING)

To generate compound word – forms, the constitutent word bases and the semantics (meanings) of compounding could be the input. The semantics can be mapped to specific relationships (functor or residual), of compounding. Thus, the object denoted by the former, latter, all. Or external word(s) may normally map to avyayibhava tatpurusa dvandva and bahuvrthi compounds. Thus, by applying the rulebase, compound words can be formed from components.

5.116 Vaidika (ACCENTED INPUT)

In Vedic processing, Samhit (continuous text or prose), Astavikrits (karma, jata, Ghana, ratha, dhavja, danda rekha, mala the eight fold combination patterns) or Pancasara Varna karma (complet characrterisation of each Vedic syllable with 26 attributes – 8 each for vowel and consonant parts and 10 for accent part) from pada patha (single words) are used as input. These processes involve accented sandhis using the concerned prtisakhay and Siksa, besides grammar rules. Thus, generation of Vedic forms does not mean taking a base for declenision in the conventional sense.

5.117 Sandhi (EUPHONIC COMBINATIONS)

This takes two or more words for input and combines two words at a time to produce the combined output form. Internal and external sandhis for vowels and consonants (using one or more rules in stages) giving mandatory and optional forms are defined.

5.12. Vyutpattih (ANALYSIS)

Word level analsyis refers to recognition of a given finished word in terms of all possible (grammatically valid) identification of base and affix combinations, This is the single most important feature of the parser for Sanskrit which would have immense utility. This could also take accented input. Thus syntactic analysis would indicate all distinct senses of a given word for further determination of intended sense under given context by semantic and pragmatic analysis. Few examples are ramah, asvah, arunach vrksye, tasya, gacchati, bhavati, yacate, matah, yanti (both noun and verb), yajeta (parasamai and Atmanepadi),vaksyati (vaha and vaca as roots) gate (all numbers), narah (rkaranta and karanta) ramah (masculine and feminine) te (8 possibilities), ma, gatum, visvasya (noun and indeclinable), asmi, aha (verb and indeclinable)etc/

The recognition of a word requires that the base and affix are available in the lexicon in the machine. Only the nominal bases are too numerous to include exhaustively and hence, an on –line lexical update facility for Pratipadikas can be provided. The program would prompt the user to exercise the option to update the lexicon with proper choice of stem, gender and paradigm type as required, when a particular word is not identified. Thusm any real text could be given to the program for analysis, without concern regarding the lexical content.

5.2. SENTENCE LEVEL

Generating sentences or analyzing them syntactically and semantically is involved in this stage. Here the individual words are to be processed in a sequence define by the input. The sentential import any introduce modifications to the sigma of individual (component) word level results and these factors are quite tricky to handle. This calls for careful choice and design of guiding principles.

5.21GENERATION

Sentence generation require details inputs. The user had to be given proper alternatives to choose from at various stages for different types of sentences. In our analysis few hundres types with well oer a thousand sample sentences collected from standard works like vaiyakarana siddanta manjusha, vaiyakaranan siddhanta kaumud sabda sakti prakasika bhatta rahasyam and sabda trangini are used as the basis. The parameters for classification could typically, be : type, accent, voice, mode, mood tense, person, number., construction, functors, beneficiary of action, verbal root, meaning, transitivity etc. Besides these, aesthetics ellipsis, stylistics etc. are also involved.

The algorithm involves selecting the activity tobe denoted, to begin with (for active type of sentence of counsel. Then to root, pre verb if required mode, voice, tense, person number, beneficiary of action, functors, their base, gender number etc. at word level are selected. Then for sentence level, construction type, accent, ellipsis, aesthetics etc are selected, the appropriate rules of generative grammar are then applied to choose the suffixes etc. for each input paratmer and the

individual words forms are generated. The sentence level requirements are then checked for compliance and the words are strung into a sentence by sentential processes (preferred order, sandhi etc).

5.22. Analysis

Here, each word is first syntactically analysed and all possible recognitions for each word are collected then, with verb as the pivot, unique functional roles are assigned for each word using akanksa and yogyata (karaka requirements and conceptual type compatibility checks) All verbs are mapped to their meaning types, karaka specification (mandatory, optional and inhibitory), and concept type compatibility criteria. Presence and capability of candidates for each necessary role of functiors in the sentence guide the process. This also helps in identifying incomplete or incompatible sentences. When all the words of the input sentence are uniquely assigned distinct functional roles, the sentential import is output by a suitable choice of sabdabodha (paraphrasing) form.

5.3. DISCOURSE LEVEL

Over and above the sentence level, pronominal and other forms of back references, anaphora, ellipsis etc. are to be particularly addressed at this stage.

5.31. GENERATION

Multiple sentences are to be generated here and an episodic structure has to be built up dynamically. The reference factors etc. are included suitably and contextual criteria enumerated in Mimamsa are utilized as also common sense issues from laukkia nyayas. This is too advanced a stage of offer any concrete points as of now.

5.32. ANALYSIS

Knowledge base of local domain and formulation of diambiguration criteria, common sense etc. are required here. In our system, a typical yajna (sacrifice) has been described as word knowledge and basic mimamsa maxims are emulated to facilitate study of the issued involved, in depth, In both sentence and discourse levels, query, processing is possible to enhance the quality of understanding the system processing. Voice change and précis output are other yardsticks to measure the extent of machine understanding. Also in analysis the result can be input to generation module for verification and vice – versa.

6.0 IMPLEMENTAION SO FAR

The various databases ad rule bases mentioned in the preceding sections have all been incorporated some to a preliminary level and some to a fairly good extent. The main point us that many of the thoughts presented here are clearly realizable as some amount of practical solution to these are available. Due to space constrains, all the details of our system are not accommodated in this paper, but are available on request. A brief information brochure on DESIKA is appended with this paper. Karaka porakaranam (Functor – Case Mapping), Vibhakti prakaranam (Sub- Functors Mapping), Dhatvartha Vislesanam (Verbal activity analysis) Padartha Vibhaga (conceptual categories) and Vakyanvaya

(sentence & Discourse Coherence) are some important features that are incorporated.

7.0. CONCLUSION

We have made a beginning in seriously studying the utility of Sastras for NLP in our system and the result in encouraging. We appeal to all Sanskritists, Linguists and Computer Scientists to support and encourage use by trying out method (and system) to enrich and revive Sastraic studies from computational angle.

REFERENCES

- "Desika – A NLU System for Sanskrit ", Project Report, Centre for Development of Advanced Computing (C-DAC), 1992.
- Devasthalik G.V. Anubandhas, of Panini Poona 1967.
- Katre, S.M. Astadhayai of Panini Roman transliteration and Eng. Tr. Delhi 1989.
- Palsule G.B. The Sanskrit Dhatupathas, A critical study Poona, 1963
- A study in compound word – forms Poona 1961
- A study in non compound word, forms Poona 1963.
- Ramanujan, P. Computer Processing of Sanskarit" Proceedings of conference on computer processing of Asian Languages (CPAL 2), IIT, Kanpur 1982.
- Computeriasation of Vedic texts. National Conference on Vedas and Shastras Tirupati. 1992.
- Mastiska yantre Sabdabodha Pratirupanam, National Conference on Vedas and Shastras Tirupati 1992.

DESIKA – NLU SYSTEM FOR SANSKRIT

GENERAL DESCRIPTION

Sanskrit is an ancient Indian Language renowned for its highly structured grammar described by sae, Panini, millennia ago. Even in modern times, the languages has evoked keen interest among linguistic and Computer Scientists for possible clues to Natural Languages Understanding (NLU) issues. DESIKA serves the need for an authentic Computer based package for its study.

DESIKA is a comprehensive package for generating and analyzing Sanskrit words. It Caters to different user communities like Academicians, students, Researchers, Linguists Computer Scientists etc.

MODULES

Generation

- Has a nominal lexicon of 3500 words selected from Amarakosa and Linganussana updatable on – lin.
- Convers ALL Nominal declension paradigm types.
- Convers ALL verbal roots of Paninis dhatupatha in all tenses. Moods (including let) verbal causative desiderative intensive, frequentative and reflesive modes, and in active and passive /impersonal voice,
- All avayaya types covered.
- Important krdantas includes.
- Output in conventional format in Devanagari script (other Indian Scripts selectable)
- Analysis
- ALL grammatically valid syntactic identifications of plain or accented, input word(non compound, currently) or sentence, output in a suitable format.
- Karaka relationships are determined based on relevant mapping rules.
- Sentence coherence based on sabdabodha theory and output accordidng to Naiyayika vaiyakarana or momamsaka formats.
- Output is paraphrased or voice change version of input in Sanskrit.

VEDIC

- Covers svara (accent) sandhis (for Taittifiya krsna Yajur Veda currently using pratisakhya and vyasa siksa)
- Generation of samhita and Veda vikrtis (combinatorial patterns) like karama, jata, Ghana etc, from Pada text as input :
- Varnakrama (exhaustive characterization of each Vedic syllable) for Pada text.

- ❖ Output in a suitable format in Devanganri script with accents marking.
- ❖ SANDHI
- ❖ Combines plain or accented words and ouputs result with the relevant rule/operation.

FEATURES

- ❖ Modules for Generation, Analysis, Vedic and Sandhi
- ❖ Serves as a PANINIAN platform for computation linguistic studies/ research.
- ❖ Substrate for development of similar packages for Indian Languages.
- ❖ useful in development of knowledge bases
- ❖ Based on ab-ultimo morphological analysis.
- ❖ Indispensable for vedic (accented sacred texts) analysis and preservation
- ❖ Suited for a tutorial on pancasandi, sadlingi karaka and Tinanta sections, (prakaranas) of Vaiyakarana siddhanta kanmudi.,

KNOWLEDGE BASE

Database is 'coded' parametically through an optimal alphanumeric code, for computational use.

Subantas	Ending , Gender, Paradigm type
Avyayas	Type meaning
Tiriantas	Roots, conjugation, indicator characters, usage, union – vowel, transitivity accents, meaning.
Krdantas	Root, suffix, semantic type
Sabdhis	Origin Internal external effort etc
Vedic accents	Udatta Anudatta svarita pracaya
Vedic Characters	Anuvara svara bhakti Rangapulta

Rulebase consists of the relevant astadhayayi sutras classified according to definition operation (including exception and extension) and interpretation. In vedic processing pratisikia and sikasa rule are also included so far, over 500 rules of Astachayayi have been covered.

SYSTEM REQUIREMENTS

- IBM compatible PCs
- GIST – 9000 add on card
- MS DOS ver 4.0 or above.

****