Prof. T. V. Gopal
Co-Ordinator, Centre for Applied Research in Indic Traditions;
Professor, Department of Computer Science and Engineering; College of Engineering, Guindy Campus
Anna University
https://www.annauniv.edu
https://vidwan.inflibnet.ac.in/profile/57545
gopal.tadepalli@gmail.com

Making Information Speak : The IT Way

Dr. T V Gopal

Abstract:

Science and Technology has been grappling with the problem of mapping accurately its lingua franca to the real world for several centuries and vice versa. One wishes that the world is in some sense a text waiting to be deciphered in a manner that all the three approaches of determining the meaning are unified. Three centuries earlier, Galileo had written, "Nature's great book is written in mathematical language".Bertrand Russell, one of the preeminent logicians of the twentieth century, once remarked of mathematical logic that it "may be defined as the subject in which we never know what we are talking about." This is the true challenging in bridging the gap between doing research and reporting the results of the research in the form of research publications.

Keywords: *Information Representation, Research, Knowledge, Print, Web*

1. Introduction:

"The problems of language here are really serious. We wish to speak in some way about the structure of the atoms... But we cannot speak about atoms in ordinary language."

– Werner Heisenberg, Physics and Philosophy, 1963

"All our reasoning is nothing but the joining and substituting of characters, whether these characters be words or symbols or pictures, ... if we could find characters or signs appropriate for expressing all our thoughts as definitely and as exactly as arithmetic expresses numbers or geometric analysis expresses lines, we could in all subjects in so far as they are amenable to reasoning accomplish what is done in Arithmetic and Geometry. For all inquiries which depend on reasoning would be performed by the transposition of characters and by a kind of calculus, which would immediately facilitate the discovery of beautiful results... Moreover, we should be able to convince the world what we should have found or concluded, since it would be easy to verify the calculation either by doing it over or by trying tests similar to that of casting out nines in arithmetic. And if someone would doubt my results, I should say to him: 'Let us calculate, Sir,' and thus be taking to pen and ink, we should soon settle the question"

- Gottfried Wilhelm Leibniz, Konzeption der Characteristica Universalis, 1677

Visualization of a concept or an idea in one's mind is the beginning of the communication. Couching this visualization using a set of chosen symbols that comprise a language has been a tough challenge. Presently, we have several organized theories on the ambience that could possibly foster ideation in an individual. Humans happen to need words in order to talk scientifically about a world that in its own right has nothing to do with language. Consequently, the determining the meaning of what is said is a complex and contested area. There are three possible approaches to resolve the issues related to meaning. They are-

- that meaning is what is intended by the author;
- that meaning is created by and contained in the text itself;
- that meaning is created by the reader.

Mathematics helps us determine the meaning of what is being communicated with minimum ambiguity and distortion. Clearly mathematics does not have the same fluency as a natural language and, even more obviously, it is rarely spoken aloud. However, mathematics is an abstract system of ordered and structured thought, existing for its own sake.

"The skeptic will say: "It may well be true that this system of equations is reasonable from a logical standpoint. But this does not prove that it corresponds to nature." You are right, dear skeptic. Experience alone can decide on truth."

- Albert Einstein, "On the Generalized Theory of Gravitation", Scientific American, Vol. 182, Pp 13–17, April 1950.

".....philosophically we are completely wrong with the approximate law. Our entire picture of the world has to be altered even though the mass changes only by a little bit. This is a very peculiar thing about the philosophy, or the ideas, behind the laws."

– Richard Feynman, The Feynman Lectures on Physics Vol. I Ch. 1: Atoms in Motion, Pearson Education, 2012.

"Data Point" emerged as a common scientific notation for presenting research results. A data point is a discrete unit of information. In a general sense, any single fact is a data point. In a statistical or analytical context, a data point is usually derived from a measurement or research and can be represented numerically and/or graphically. The term data point is roughly equivalent to datum, the singular form of data. Unfortunately, very few text-books ever bother to define the term 'data'.

'Data' is the plural of 'datum'. A dictionary definition of 'datum' is "any fact assumed to be a matter of direct observation ". This is inadequate for machine generated random numbers. 'Data' is any symbol, sign or measure which is in a form which can be directly captured by a person or a machine.

The DIKW model or DIKW pyramid is used to explain the ways to move from Data (D) to Information (I), Knowledge (K) and Wisdom (W) with a component of actions and decisions. The questions Who? What? When" and Where? on Data result in Information. The question How? On Information results in Knowledge. The question Why? on Knowledge results in Wisdom.

"Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?" *– T.S. Eliot, The Rock, 1934*

A research paper should reflect the DIKW model without any loss of generality of the methodology adopted by the researcher. The transitions innate to the DIKW model clearly knit the parts into the whole through establishing the connections with Information and Aggregates through Knowledge.

This endeavor of documentation by the researcher enables the path -

"Researching -> Absorbing -> Doing -> Interacting -> Reflecting.

It is thus the duty of the researcher to report the research to enable the chosen field of research thrive and facilitate growth by sharing the knowledge with fellow researchers.

2. Writing for Print and Web:

2.1 The basic characteristics of a web user are-

- 79% scan - They DO NOT Read. Print is for reading and understanding.
- Reading from the computer screen is tiring and 25% slower than reading from the paper.
- Users of the Web are highly impatient.
- Each page competes with millions of similar ones. The content must catch the attention of a web user in a few seconds.
- Life is hectic. No web user has more than a few minutes to spend on the content and grasp the gist of the content.

2.2 Writing for the web user is a special skill that flourishes with the information about-

- Average Age
- Gender
- Major Occupation
- Marital Status
- Monitor Screen Sizes
- Computing Platform
- Browser Used
- Screen Resolution
- Average Age
- Gender
- Major Occupation
- Marital Status
- Monitor Screen Sizes
- Computing Platform

- Browser Used
- Screen Resolution

The content to be hosted on the web needs recasting if it has to go beyond improving the accessibility to the research results. Please see the Figure 1 below.

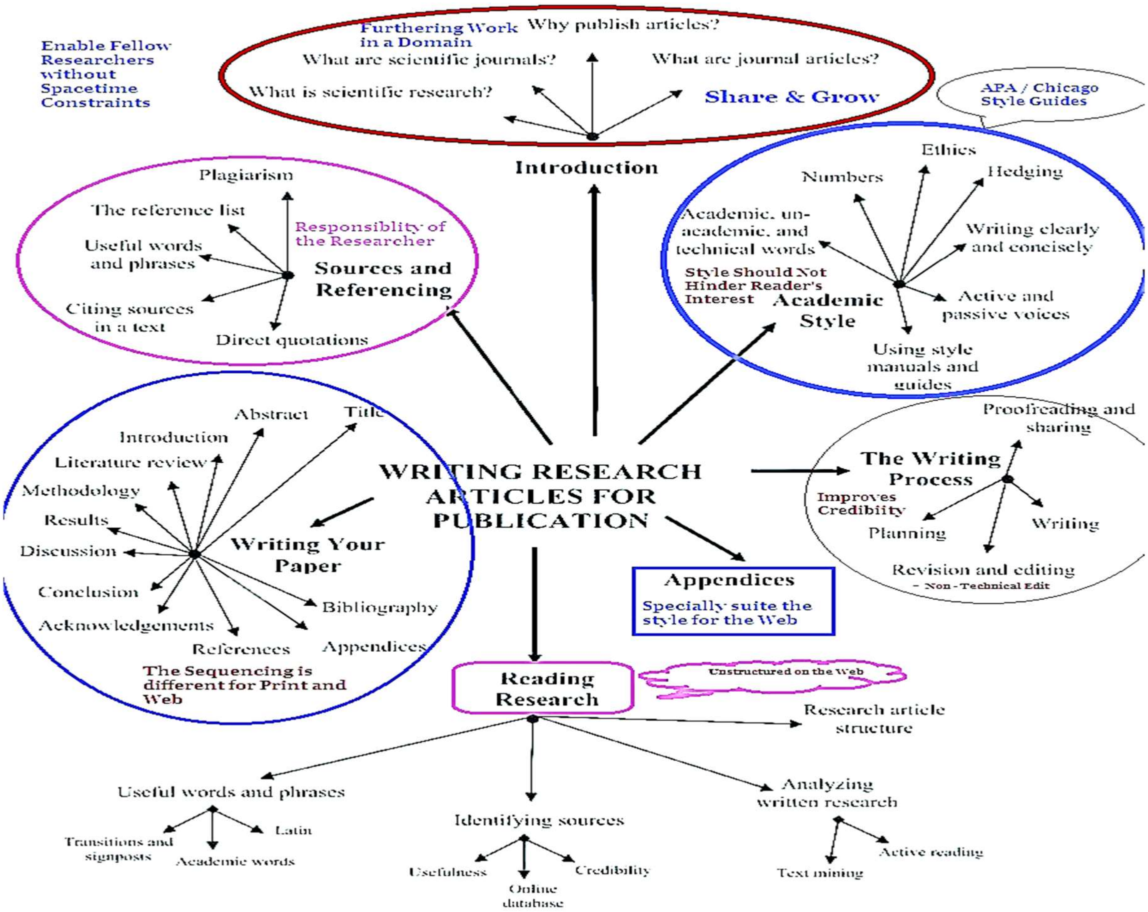

Figure 1: Writing Research Papers for reading in Print and on the Web

There have been many web site usability studies. Keevil Usability Index uses a simplified checklist approach. The items in the checklist have a dichotomous scale ["Yes" or "No" response], and the total score is just the sum of the checked items. The Keevil Usability Index checklist consists of 203 questions arranged in the following five categories-

i) Finding the information: Can you find what you want?
ii) Understanding the information: After you find the information, can you understand it?
iii) Supporting user tasks: Does the information help you perform a task?
iv) Evaluating the technical accuracy: Is the technical information complete?

v) Presenting the information: Does the information look like a quality product?

This is a simple method of checking the suitability of the research paper for reading on the web. Writing for the Web requires a different structuring to the research paper. Research results presented in the form of tables, charts, graphs are suitable both for the print and the web versions of the paper. Choosing the presentation for the DIKW model makes the real difference.

3. Hints for the Choice of Presentation:

- A table can present data more concisely than text can, and it is more accurate than graphic presentations. The conventional sentence is a poor way to show more than two numbers because it prevents comparisons within the data.
- A table facilitates comparisons among data because of the arrangement of the data into rows and columns. "Tables are preferable to graphics for many small data sets. … Tables also work well when the data presentation requires many localized comparisons." - Tufte, E. R, "The Visual Display of Quantitative Information, Cheshire, CT: Graphics Press, 1983.
- The table should present meaningful data. The data should be unambiguous. The table should convey ideas about the data efficiently. Ideally NOT more than 10 Rows.
- Overall trends about the information, however, are more easily seen in charts and graphs.
- Charts are graphical representations of an underlying statistical table.
- As a general rule, a chart or a graph will be a more effective way of presenting larger amounts of data, although with less accuracy, than a table.
- With large amounts of data, charts often are more effective at revealing trends and relationships among the data.
- Presentation of data in a visually appealing fashion, without sacrificing the richness of the data, is an art.
- Data should be presented in a manner that will communicate the maximum information in the most efficient manner.
- Pictorial presentations are advantageous because even a novice without any technical expertise can assimilate the information by looking at them.
- Scatter graphs are widely used in science to present measurements on two (or more) variables that are thought to be related;
- The values of the variables as the y (vertical) axis are thought to be dependent on the values of the variable plotted along the x (horizontal) axis. The latter is said to be the independent variable.
- A line that displays a trend is called a trend line or line of best fit. This line is obtained by using a technique called Regression.

- The pie chart is a staple form of data presentation graph. Used properly, it can be an effective way of presenting a small number of pieces of data. The pie chart should be used only where the values have a constant sum (usually 100%). It should be used where the individual values show significant variations; a pie chart of seven equal values is of no use. It is often worthwhile adding annotations, especially the values for each category (thus saving the need for a separate table of data values). It should be used when the number of categories (`slices') is reasonably small; as a rule of thumb the number of categories should be normally between 3 and 10.
- The column (vertical) chart should be used for data which have some `natural sequence' in the categories.
- The bar (horizontal) chart should be used for data with no natural `sequence' in categories.
- Line charts are similar in some ways to scatter graphs, with the extra constraint that the values of the independent (x) variable have their own sequence. Moreover, those values are a sample from a (presumed) continuous series, such as temperature, pressure or commodity prices.
- Some specialized domains need special charts such as Area Charts, Polar Charts, Triangular Charts and Time Series Charts.
- The tables, charts and graphs may be made attractive and easy to read by judicious use of Color, Point shape, Fill pattern and Line style.

4. The following guidelines are very helpful in choosing the representation of the research results:

- Communicate relevant information explicitly OR What needs to be conveyed.
- Different graphic formats highlight different information about the same complex data. Line graphs are used for conveying trends (graphic convention is to only use for continuous data, but this convention is violated that to make my point sometimes)
- Bar graphs to make categorical comparisons
- 3-D graphs if exact data isn't important, but a complex relationship or some idiosyncratic data points are important.
- Pie charts are not good for exact values, good for comparing proportions.
- Avoid forcing the reader to make any "calculations" or complex inferences.
- Adding 3-D is not always so good. One can lose ability to read exact data points, and one can occlude important data.
- Lots of colors and designs are not necessarily helpful
- The lines are to be labelled and using legends may be avoided

➢ The symbols or lines distinct if they are to be compared. This is essential if data are complex.

Modern software makes it far too easy and tempting to add a third dimension to objects in graphs. This functionality is thrown in because people expect it, not because it's useful. It is far better to impress your readers with graphs they can easily understand and use, rather than graphs that look like a video game and are difficult to interpret. There are many commercial and open-source graphic software.

There are several important diagram formats that are less obviously 'number-oriented' than those we have considered so far. Some may incorporate quantities as a part of their overall structure, but most present qualitative information. However, their wide usage in engineering indicates that knowledge of their characteristics is important for the practicing engineer. Some widely used structure charts are Flow Charts, Organizational Chart and Gantt Chart.

5. Conclusions:

Print is linear, author –driven narration. The web is non-linear, reader – driven and is ruthless pursuit of actionable content. Making the information speak the IT way is common to both Print and Web. Understanding that the web is a three-ring circus of Information, Exchange and Entertainment is essential for the research and researcher to excel.

References:

- Brient C Oberg, "Interpersonal Communication", Jaico Publication House, Mumbai, 2005.
- David Peat F, "Mathematics and the Language of Nature", Mathematics and Sciences, Word Scientific, 1990
- Gelman, A., Wainer, H., Briggs, W., Friendly, M., Kwan, E., & Wills, G, "Why Tables Are Really Much Better Than Graphs [with Comments and Rejoinder]. Journal of Computational and Graphical Statistics, Vol. 20, No. 1, Pp 3-40, 2011.
- Keevil, B, "Measuring the usability index of your website", Proceedings of the Sixteenth Annual Conference ACM International Conference of Computer Documentation, Quebec, Canada, 24-26 September 1998, Pp. 271-277.
- Nathan, R.J., Yeow, P.H.P. and Murugesan, S, ."Key usability factors of service-oriented web sites for students: an empirical study", Online Information Review, Vol. 32 No. 3, pp. 302-324, 2008.
- Ronald B Adler and Jeanne Marquardt Elmhorst, "Communicating at Work", McGraw Hill, 2002
- Stephen C Few, "Show Me the Numbers – Designing Tables and Graphs to Enlighten", Analytics Press, California, USA, 2004.
- Steve Talbott, "The Language of Nature", The New Atlantis, Winter 2007