Understanding Information

1. Data, information and knowledge
The terms data, information and knowledge are used interchangeably, and there’s little consensus on what each precisely refers to. However, here are some preliminary ideas.

**Data:** Data is raw material – numbers and strings of letters with no precise context or meaning. A list of numbers – 26, 24, 27, 26, 28.5 – is a good example of data.

**Information:** Information is defined in a number of ways. Some people think of information as ‘processed data’. Other popular ‘definitions’ are listed below:

- ‘Data endowed with relevance and purpose’. (Peter Drucker – try and find out who he is)
- ‘Data becomes information when its creator adds meaning’. (Davenport and Prusak)
- ‘A representation of reality’. (Jessup and Valacich)
- ‘An organized, meaningful and useful interpretation of data’. (Senn)

Keith Devlin, the author of the recent book InfoSense, has a variety of definitions for information. Loosely speaking, one of them is that information is a representation of data combined with a set of regularities and conventions that one uses to encode and decode this data.

**Knowledge:** Again, there are a multitude of definitions for knowledge.

- ‘Some form of accumulated information – a body of guidelines and rules that are used to manipulate data in order to make it suitable for a given task’. (Jessup and Valacich).
- ‘An awareness and understanding of a set of information and how that information can be put to the best use’. (Senn)
- ‘A fluid mix of framed experiences, values, contextual information, and expert insight, that provides a framework for evaluating and incorporating new experiences and information’. (Davenport and Prusak)

Keith Devlin simplifies this into the equation:

\[ \text{Knowledge} = \text{Internalized information} + \text{Ability to utilize this information} \]

So much for definitions. Think about these ideas. We’ll discuss them in class, and work towards a common understanding of these terms.

2. Types of information
There are five types of information that one deals with in the context of information systems:
Numbers: Raw, numerical facts of a situation. Some examples include the closing stock price of Cisco on September 1st, the average number of points scored by Patrick Ewing per playoff game, the number of students enrolled for our course and your score on the midterm exam.

Text: A set of strings of characters. Text is also called narrative information. The stuff you are reading right now is textual information (and it’s knowledge of the highest form, of course 😄).

Sound: Audio information – a song, a speech by the President, the noise made by a car’s brakes. Over the last fifteen years, audio information has become closely associated with information systems. For instance, the CD’s you own have the music on them represented in the same way that a computer represents all its information.

Images: Information in visual form. An image can summarize numerical information, or can simply represent a drawing or a photograph. The visual representation of numerical information is very important, since it seems that human beings understand visual information better than tables of numbers.

Video: A timed sequence of images, superimposed with audio information. There’s a lot of video information on the Excel CD’s you’ll be using in this course.

3. How computers represent information

Computers are electronic machines that can perform precise instructions extremely fast. The important observation here is that they are electronic – eventually, they represent all their information as a collection states of binary switches – switches that can be either on or off, and whose state is either ‘on’ or ‘off’. It’s useful for us to think of these states as 1’s and 0’s, which is why we say that eventually, computers represent all information as sets of 1’s and 0’s.

There is a system of mathematics called binary arithmetic, which uses only two digits – 1 and 0 (analogously, our favorite system, the decimal system uses ten digits). Therefore, we treat the representation of information in computers as representations in binary. A binary digit is called a bit, and a set of eight bits is called a byte.

It’s fairly easy to conceptualize representing numerical information in binary. However, the other forms of information we described can also be (and are) represented in binary:

Text: Each character on your keyboard has a unique binary representation; for instance, A is represented as 100 0001. The most common code that maps the characters to their binary equivalents is called ASCII (American Standard Code for Information Interchange)

Sound: Audio information can be approximated as a timed sequence of audio frequencies (how high or low the audio is) and amplitudes (how loud the sound is). Therefore, a simple way to represent sound in binary would be as a sequence of pairs of numbers (one per millisecond, say), the first representing the frequency, and the next representing the amplitude. If this is hard to understand, don’t worry; just skip it.
Images: Images can be broken down into a large grid of spots, called pixels. Each pixel has a color; if we represent each color as a unique multiple-bit binary number (common lengths are 8, 16, and 32), then this grid of binary numbers represents the image. Again, if this is hard to understand, just skip it.

The details of how to represent images, sound and video on computers are of deep interest to engineers – for our purposes, it’s sufficient if we understand that it is possible to represent these images, sound and video in the same ‘language’ that computers use to represent numbers and text – as sets of binary digits.

4. Properties of information

Information has a number of properties that distinguish it from physical goods like televisions and Coke cans. Since information has become an important ‘good’ in our economy, it is helpful to examine a couple of these properties:

Infinitely replicable: If you have one copy of a particular piece of information, it is possible to make as many copies of it as you want. This is in stark contrast to a physical good, like a Coke can (tragic, isn’t it). Also, if the original copy of the information was valuable, then it is possible that each new copy is equally valuable – introducing a whole host of reselling problems, if one is selling information.

High fixed costs, almost zero marginal costs: In other words, information is typically costly to produce, but very cheap to replicate. Think about, for instance, the recent Star Wars movie. It cost hundreds of millions of dollars to produce (high fixed cost) – but the entire movie will soon be put onto a DVD, and replicated for a few cents a copy (almost zero marginal cost).

Experiential: You may not know the value of a particular set of information until you actually ‘experience’ it. For instance, you don’t really know how valuable the lead story in the Wall Street Journal is going to be today until you’ve read it, and you need to buy it in order to read it. Also, there’s only so much information one can experience in a given time period, since we have limited attention spans, and our brains have limited processing capabilities.

These properties of information make its economics very interesting, and very different from the traditional economics of physical goods. We’ll explore this subject later in the course.

5. The information value chain

There are a number of ways in which one can commercially generate value from information; five of them are listed below. If one actually examines the businesses of companies in the Internet economy, one will find a number of them rely on simply adding value to information in one or more of the following ways: aggregating information, organizing information for easy retrieval, filtering out relevant information based on a customer’s needs, packaging information in a format that makes it easy for the customer to process it, and distributing information to a large group of customers. More on this in class.

---

1 These are loosely based on the value matrix discussed in the article “Exploiting the Virtual Value Chain”, by Jeffrey Rayport and John Sviokla of Harvard Business School.