Appendix

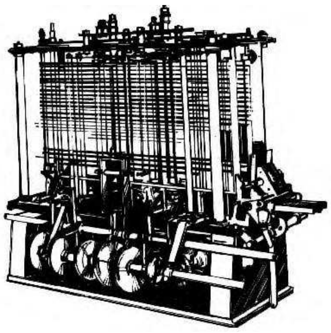


Figure A.1: The Analytical Engine developed by Charles Babbage is regarded as the first programmable calculating machine. (check permission)

A.1. Information Theory

We have defined information as something which changes the behavior of a system which receives it. It is difficult to specify exactly what those critical factors will be but to the extent that we can specify them, we may be able to figure out how to transit them. Information can also be defined as selecting one alternative from among several others. Transmission of representations though that isn't always transmission of rich information.

If we can figure out what needs to be transmitted, we can determine the number of bits required to transmit them optimally. Examples of the surprising-ness of information. Being notified that you have won the lottery is truly surprising since the chance of that is quite small. There is a wide range of applications for Information Theory. Though, it is difficult to understand how much information is being transmitted without knowing how the information is represented.

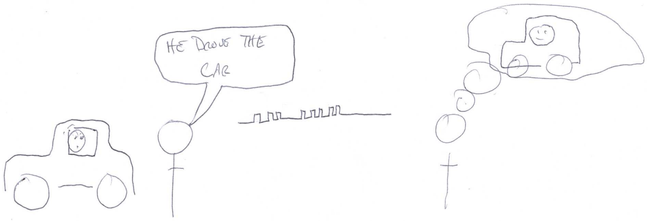


Figure A.2: At least representations can be coded and transmitted in terms of bits. If they can be unpacked as they were encoded, then there can be perfect information transfer. But, of course, not two people will not have the same encoding and decoding systems.

A.1.1. Measuring Information: Entropy

For communication systems, it is desirable to encode as much information as possible into a narrow channel. This was the basis of the simple information transfer model of communication. We want to determine the most compact representation for a message. This is useful for instance, for data compression, data storage, and for Hidden-Markov Models (11.3.3, -A.5.5). Given a vocabulary, we can calculate the fewest number of bits needed to transmit a message[68].

In complex environments, is it really possible to measure information?

Suppose we had a group of four people and we had to pick one of them (Fig. -A.3). Assuming they are equally likely to be picked, the probability would be 0.25. We would use $Code_1$ or $Code_2$ and we would know that $Code_2$ is optimized. Since the log is usually $base_2$, information is usually measured in bits. Indeed, the word “bit” is derived from the phrase “binary information unit”.

$Distribution_1$			
Person	**Probability**	$Code_1$	$Code_2$
Abu	0.25	1000	00
Bob	0.25	0100	01
Cathy	0.25	0010	10
Dwayne	0.25	0001	11

Figure A.3: Two coding systems for identifying which of four individuals might be selected in a lottery.

The self-information of a message is related to probability of that message; that is, how likely or predictable that message is (Eq. -A.1).

$$I(m) = -logP(m) \tag{-A.1}$$

“Entropy” is a measure of the disorder of a system o set of messages (Eq. -A.2). For the data in Fig. -A.3, the entropy is $H = X$. Because each person is equally likely to be picked, we cannot do any better than chance in guessing who that person is. However, this also means that the codes we use to identify the person can be very efficient. For the probabilities in $Distribution_2$ (Fig. -A.4), the entropy is $H = 1.8$ and the codes to indicate which of them has been selected are not as efficient as those for $Distribution_1$.

$$H(X) = -\sum_{i=1}^{k} P(x_i)log_2P(x_i) \tag{-A.2}$$

$Distribution_2$	
Person	**Probability**
Abu	0.40
Bob	0.15
Cathy	0.10
Dwayne	0.35

Figure A.4: Unequal probabilities of being selected, as shown here, have lower entropy than equal probabilities (shown in Fig. A.3).

Perplexity.

Another way to think of entropy is as an indicator of the average “surprise” of the choices. When the probabilities of all choices are equal, as in $Distribution_1$, the level of surprise is maximized. Another way to look at this is ask what is the additional value contributed by a given source of information. Maximum entropy. Knowledge at the receiver’s end can compress information much more.

$$H(X,Y) = \tag{-A.3}$$

Mutual information.

Information valuation. Bayesian models for deciding how much to value information sources.

A.1.2. Communication Channels

Once we have a measure of information, we can compare the amount of information able to be transmitted on different channels (Fig. -A.5). Communication models ((sec:communicationmdoels)). We might ask how much information can be transmitted with a fixed number of bits in a communication channel. The bits able to be transmitted per unit of time, is the channel capacity which is also known as the “bandwidth”.

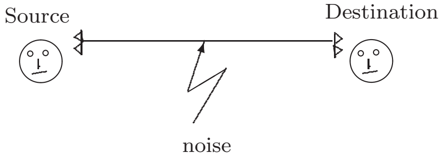


Figure A.5: Information transfer (adapted from[68]). (check permission)

It is possible to calculate bits of information based on assumptions about the receiver’s capabilities. If the communication channel is imperfect (e.g., noisy), we could calculate how much information can be transmitted. Signal processing equations can be used to support tasks such as speech-processing (11.3.3) and evaluating the quality of machine translation (10.13.1). Specifically, we can model translation as a noisy channel between a source and a receiver[9].

A.1.3. Applications of Information Theory

Applications Sensor networks.

Can we really measure information and meaning in people’s heads? Can we even usefully measure how much information there is in a complex information such as a book or a videotape by measuring the number of bits in a digital copy of that resource?

The Redundancy of Natural Language

Natural language is highly redundant. Put another way, every letter, phoneme, or word is not totally surprising. You should be able to make a good guess about the missing word in the sentence: “You are reading a book about Information ________”. Redundancy in natural language prevents misunderstanding. the amount of redundancy in natural language is related to perplexity. Can we estimate the amount of redundancy.

Language is, effectively, a coding system. Several approximations to English are shown in Fig. -A.6. These approximations are based on the likelihood of letter and word combinations. One application is to machine language translation (10.13.1).

Level	**Example Approximation**
First order Word level	REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.
Second order Word level	THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

Figure A.6: N-gram approximations to English[67]. (check permission)

N-grams are particularly useful for speech processing where the sequences of phonemes is highly predictive of specific words.

Example = Morse code

A.2. Compression

The content in almost all information resources is redundant and can be compressed. Most often this is done to reduce the amount of material which has to be transmitted or stored. Compression is a type of representation, but it is generally not a representation from which semantics can be easily extracted. Information theory (-A.1.0) can measure the effectiveness of the compression.

A.2.1. Issues for Compression

What Makes a Good Compression Scheme?

There are many options in compression; as with other aspects of information systems, the selection of these options depends on the requirements of the users. There is a close connection between compression and the preservation of meaning.

Efficiency Two main types of compression are termed lossless and lossy. If it is possible to get an exact copy of an image back after compression, then the compression is “lossless”. If something is lost so that it is not possible to retrieve an exact copy, then the compression is “lossy”. Most compression algorithms are lossy.

Codes and compression. Fixed code length versus variable code length.

We can measure compression by the “compression ratio,” which is the ratio of the size of the file before compression to its size after compression.

Codebook.

Adaptive compressions. Self-correcting codes.

Content Dependencies Some compression schemes are good for specific content. GIF compression works particularly well for line drawings while JPEG compressions are especially good for full-color images. Compression needs for astronomy or medical images are very different from those for video games.

The amount of variability in content will affect compression needs.

Self-referential codes.

To the extent that a compression scheme captures semantically meaningful events.

some of the semantics can be extracted from the compressed formats. Therefore, the compressed representation may also be useful for retrieval.

Delivery, Storage, and Decompression Robust to packet loss[??]

Some storage devices (e.g., CDROM) and some networks (e.g., modems on voice telephone networks) deliver fixed data rates. Other systems deliver variable-bit rates (VBR).

A system can provide real-time delivery, or may be real-time interactive.

Layers of multimedia are prioritized. Compression matched with priority for transmission.

The recipient has to know how to decompress the message.

Compression and decompression take up a certain amount of computational resources; the right approach can optimize results.

One might opt for software compression or hardware compression.

Tradeoffs are made regarding, for example, the amount of disk space used versus the speed of searches.

"Transcoding" is the transfer from one compression system to another one. However, there can be a substantial loss of quality in the process.

Two Paradigms for Compression

Compression may be thought of as a type of representation (1.1.2). An optimal compression would be based on human perception and information processing, but algorithmic compression may not seem to be based on the semantics of the material being compressed. The compressed signals necessarily follow the characteristics of the content. Speech signals in a telephone are based on the range of speech necessary for speech comprehension. Because compression often causes loss of content, the issue is how to minimize that loss in comprehensibility. These differences can be understood in terms of Information Theory (-A.1.0).

A.2.2. Text Coding and Compression

Compression of the message. Text compression tends to emphasize lossless compression techniques because there is relatively little data and any loss can be significant. However, there may not be too much need for this since is it cheap to transmit text.

Many coding systems have been developed. Huffman Coding and Huffman Trees. Earlier, we compared the entropy of coding two sets of events with a two-bit code (-A.1.0). Huffman codes attempt to match the length of a code with the frequency of its occurrence in the family of messages to be transmitted. If we are transmitting letters, given that the letter "e" is the most common letter in English, it would be the shortest term (Fig. -A.7).

A second version of this can be seen in Fig. ??. Letter and frequency (Korfage example).

$Distribution_3$		
letter	**Probability**	$Code_1$
e	0.675	1
i	0.125	01
o	0.125	10
l	0.125	11

Figure A.7: An example of a Huffman Code in which high probability items are given short codes. (check values)

A.2.3. Image Processing and Compression

Coding — compression - formats. Pictorial material comes in many forms and the optimal coding for those varying schemes can be very different. A black-and-white line drawing will have very different compression characteristics from those of a complex colored photograph. Bitmaps of text as for OCR. [??] Displays and printing technology are described in -A.18.2.

Edge detection, shape detection, textures.

In run-length encoding, the sequence B,B,B,B,B,B,A,A,A could be encoded as 6B3A That is six repetitions of B followed by three repetitions of A.

The technology for handling still images is now fairly well established. They are easy to digitize, compress, transmit, and embed in documents.

Visual words.

Object detection techniques (11.2.2) are similar to those used for still images.

Digital Encoding of Images

In a black-and-white image, only the brightness of pixels is measured. Gray scale. Color depth (Fig. -A.8).

Pixels			
R	**G**	**B**	**Visible Color**
0	0	0	Black
256	0	0	Red
0	256	0	Green
0	0	256	Blue
256	256	256	White

Figure A.8: All possible colors can be coded as levels of Red (R), Green (G), and Blue (B). 8 bits is often used for each coding each of these base colors allowing 256 shades of each one.

As noted earlier (4.2.3), human color perception is complex. A variety of systems have been developed for coding colors. Some are based on the human perceptual system (such as HSL) and some are based on technological convenience (such as RGB). The most common system is RGB (red, green, and blue). The HSL system deals with hue, saturation, and luminance; some claim that it is closer to the human visual system as explained in 4.2.3, above. Still another system is YIV; Y stands for luminance, I for the red-cyan dimension, and V for the green-magenta dimension. YUV is used for broadcast television; here, Y = luminance, U = blue-Y, V = red-Y.

One element in color coding is the way the colors are distributed on the color space; another element, color depth, refers to the number of bits allocated for the representation of each color. One common system uses one byte (8 bits) assigned to each of the red, green, and blue channels. This allows encoding of 256^3 (65K) colors.

Image Processing of Pixels

This kind of processing is not object-based. The quality of the image can be improved by adapting pixels. From pixels to image processing.

Some specialized processing is model-based.

Noise suppression.

Get a signal![??]

Dithering

Image recognition.

Content based image retrieval. CBIR.

Image Compression

Image compression reduces the amount of data necessary to reproduce images. This facilitates storing data on a disk or sending it over a network. Ideally, we could find a small set of data and a few simple parameters that describe the complexity of an image. There is a great deal of redundancy in most images and this redundancy can be used in many ways to compress the image. Adjacent pixels are often similar in color. This can be used to take advantage of lossless compression with run-length encoding similar to that described for text (-A.2.3). TIFF compression is lossless.

Lempel-Ziv-Welch (LZW) GIF images use the Lempel-Ziv-Welch (LZW) algorithm which is based on probability functions.

Discrete Cosine Transformation (DCT) Discrete Cosine Transformation (DCT) converts the colors of the image to frequencies. DCT is like Fourier transformations (-A.2.4). Low frequencies encode the dominant colors and higher frequencies encode the transitions.

Wavelets Wavelets are similar to DCT in characterizing an image on its frequencies. They are also similar to Fourier compression (-A.2.4). However, DWT Wavelets are more flexible in representing objects than are the trigonometric functions in DCT.

Fractal Compression Fractal compression uses repeated application of an algorithm to approximate the original image (-A.10.2), which is generated by recursive application of the program.

Transformations: rotation, dilation, reflection

Image Formats

Beyond the compression algorithm applied to individual frames, a wide range of compressed image formats has been developed, a few of which are in widespread use. JPEG and GIF are the most common formats and will be considered here. Graphic Interchange Format (GIF) uses the LZW algorithm. The GIF specification includes composite images. These can be used to create apparent motion in an image and are known as animated GIFs.

The JPEG (Joint Picture Experts Group) format is blocky.

Typical compression ratios for JPEG are on the order of .[??] A 100KB file might be reduced to 10KB. There are several levels of quality for JPEG images and the quality selected will affect the amount of compression. JPEG is generally better than GIF for color pictures because the underlying DCT transformation allows a wider variety of transitions to be represented.

Transmission Progressive transmission allows displays of varying qualities as they are received across a network. Thus, a partial version of the image can be displayed before the transmission is completed. This sometimes works as interlacing.

Example: JPEG-1

8x8 pixel blocks. Slices,[73] has a detailed discussion of the JPEG standard and a good overview of other compression techniques. Discrete Cosine Transform (DCT, as described below). Quantized Q-matrix.

Example: JPEG-2

Object-based[??]

Scene Recognition

A.2.4. Audio Processing, Compression, and Coding

Audio Coding and Compression Algorithms

The choice of the coding algorithm depends on what is being encoded and the environment in which the it has to operate. The two most important applications for audio are speech and music. There are large differences in the encoding requirements between music and speech. Speech has a relatively narrow dynamic range while music may vary to a much greater extent. Some codes must operate in environments where some of the data is lost during transmission.

Sound waves are converted to analog electrical signals by a microphone. To create digital audio, these analog signals must be converted to numeric values. There must be an analog-to-digital conversion (AtoD). AtoD conversion is also known as Pulse Code Modulation (PCM); it involves two steps: sampling and quantization. Sampling is the number of times a signal is coded per second. To get a full coding of a signal, it must be sampled at twice its frequency. Once the signal is sampled, it must be assigned a numeric value which can be represented in a computer word. The code is usually linear, but can also be logarithmic.

Quantized digitized audio on a CD-ROM is not compressed; the quantized samples are just stored as they are coded. This simplifies the electronics and there is no need to control the rate of playback. When storage capacity is at a premium or network congestion is a problem, compression greatly reduces the amount of data to be stored. In contrast to compression, in which the number of bits is constant regardless of what is contained in the audio file, Variable Bit-Rate (VBR) coding uses total bits when

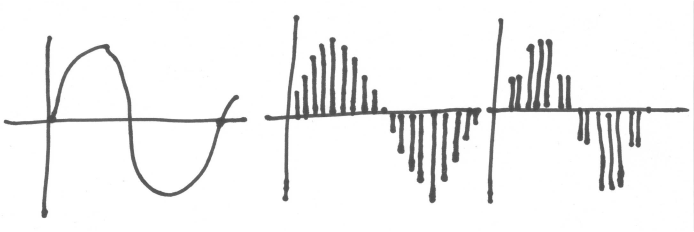


Figure A.9: An analog sound wave (left) can be digitized (center) and then it can be "quantized" to a limited number of levels (right).

the data are easily compressed.

As noted earlier, speech contains phonemes which appear as formants (11.3.3), relatively regular bursts of sound that are related to the meaning in the words. Linear predictive coding (LPC) estimates the pattern of these formants and codes them efficiently for transmission (Fig. **??**). These codes can be used to determine filters. A variety of systems have been developed to make LPC coding more efficient. For instance, DPCM (Differential PCM) encodes the differences between pulse code samples. If the tones are steady, then little additional information needs to be transmitted. DPCM is analogous to frame differences for video (-A.2.5).

Audio Processing

Fourier Analysis The French mathematician Joseph Fourier had the insight that complex waves could be described as a combination of regular sine waves. Because each sine wave has a known frequency, a Fourier analysis of speech shows the main frequencies in that speech. Fig. **??** shows the decomposition of a signal by Fourier analysis.

This is the analysis described in the spectrogram in Fig. 11.15. The frequency of speech changes rapidly as the person produces different sounds. Characterizing these changes in frequency is important for speech processing (11.3.3). A particularly useful function for determining the spectrum is the Fast Fourier Transform (FFT).

Compressed Audio Formats Beyond the specific codec used, data may be formatted so that it can be stored and transmitted. A CD has no compression; the physical design of a CD is described in -A.20.1.

Some common audio formats are U-law, WAV, and AIFF. While MPEG-2 is primarily a video standard, the MPEG-2 audio standard has been adopted for studio quality sound reproduction. It has 64kbits/s per channel with five main channels (left, center, right, and 2 for surround sound), and other specialized channels, such as one for low frequencies.

Secure Digital Music Initiative (SDMI)[??]

Specifically, MP3 is audio layer 3 of the MPEG2 standard. The popular MP3 music standard is part of MPEG2.

A.2.5. Video Processing, Compression, and Coding

Video requires far more data than audio; therefore, compression is particularly important for networking and storage. Additional discussion of video networking and video displays is found in (-A.18.2).

Frame Differences

In a video, one frame is much like the next. This means that they do not have to be presented separately. Frame differences in video often reflect the motion of objects. Because frame differences are widely used in compression algorithms, it is often possible to detect motions from compressed video. The top panels of Fig. -A.10 show an object moving from left to right. The middle right panel shows the overlap of the two positions of the object, and the lower right panel shows the frame differences.

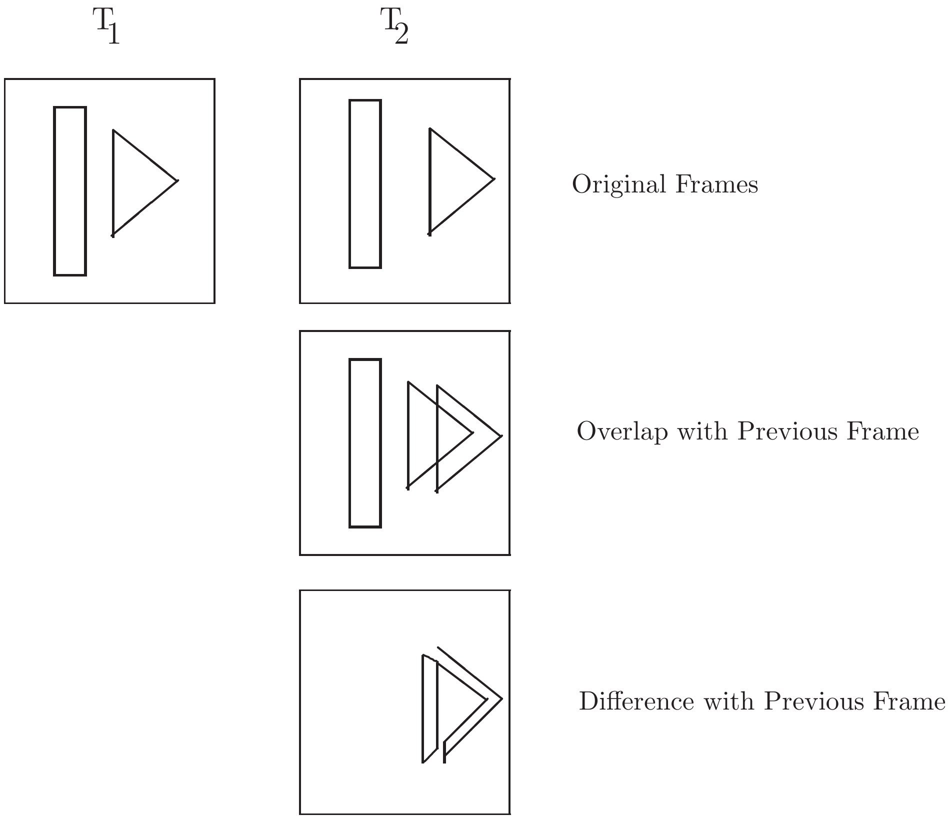


Figure A.10: Consider a two-part object that moves as shown in the top two panels T_1 and T_2. If all Panel T_1 is transmitted by video, then all of T_2 does not need to be transmitted; only the differences between the frames need to be sent. The panel in the middle row shows the two frames superimposed on each other and the bottom panel shows the differences in the two frames. Panel T_2 can be generated from T_1 by applying that difference. For this example, the area in the open trapezoid should be reset to white while the black area needs to filled in (lower panel). (smaller)

Digital Video Compression Algorithms

For digital video, the principles of color coding are similar to those for still images (-A.2.3), with the addition of a temporal dimension. Codecs for the compression and decompression of audio were described earlier; video is also compressed and decompressed by codecs.

The effectiveness of the codec depends on the context in which it is being used. If there is a lot of action in a video clip, then fresh frames may be most effective, but if the clip is just a head-and-neck-shot of a person talking, then frame-differences should be sufficient.

Forward DCT Only those pixels that change from frame to frame need to be updated. This allows only the differences to be transmitted rather than all the pixels for every frame. There may be drift from the original picture and the image will need to be refreshed by re-sending the entire current frame. This fresh frame is called a “key frame” (in MPEG, they are known as I frames).

DCT DCT is like still-image encoding. Entropy coding (-A.1.1).

3-D Fractal Video Compression Also combines with bin-trees.

Adaptive Algorithms As the name suggests, an adaptive algorithm adjusts to the type of content which is being compressed, and hence involves more content-specific coding. MPEG-4 ((sec:mpeg4)).

Digital Video

This section considers several different formats for digital video.

Version	Brief Description	Section
MPEG-1	1.5 MB/s video (PC quality)	This section
MPEG-2	Studio quality video (45MB/s)	This section
MPEG-4	Component descriptions	This section
MPEG-7	Video content description	11.6.2
MPEG-21	Framework for services	7.8.4
MPEG-A		
MPEG-V		

Figure A.11: Summary of MPEG standards.

The MPEG-1 standard is for PC-quality video (less than 1.5MB/s). Fig. -A.13 shows an EG1 stream of frames and frame differences. The standard specifies I, P, and B frames. The I frames are “key frames”; they are essentially JPEG images. The B and P frames are obtained from frame differences (Fig. -A.12). Decoding in software is practical; encoding is computationally expensive and often is not done in real-time.

Frame Type	Description
I frames	They are JPEG (-A.2.3) images and are high-quality reference frames. Transmission of these requires channel capacity. On some systems, these are called key frames.
P frames	forward compression
B frames	use bi-directional (both forward and backward) compression. These are particularly difficult to do in real time.

Figure A.12: Types of MPEG-1 frames as shown in Fig. A.13.

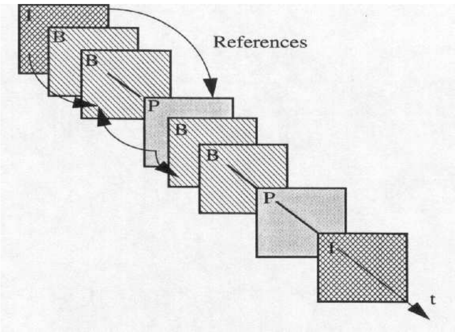


Figure A.13: Mix of frames in an MPEG-1 stream. (redraw-K) (check permission)

A.3. Graph Theory

We have seen many examples of graphs. Graphs are composed of two types of objects: nodes and links. We have seen applications of graphs across many of the topics in this book such as characterizing hypertext (2.6.3) and social networks (5.1.0). Along with state machines and grammars, graph theory is a part of a field called discrete math.

Facebook equation. Tie strength.

A.3.1. Types of Graphs

The macroscopic structure of the graph can become important for large graphs. Different types of graphs have different properties. When a number of nodes are connected by links, the pattern of the connections may be characterized. Links in graphs may be directed, that is, they allow connections in one direction but not in the other. The connections of pages on the Web form a "directed graph". If it is possible to get back to a node by some route once it has been left, then the graph is said to have cycles. If there are no cycles in directed graphs, they are said to be "acyclic" and the full graph is said to be a "directed acyclic graph" (DAG) (Fig. -A.15). Citation networks, for instance, are DAGs – time flows in only one direction.

Graph	Any set of connected nodes.
Lattice	
Directed graph	Edges have a direction.
Tree	Trees have only one path connecting any two nodes.

Figure A.14: Types of graphs.

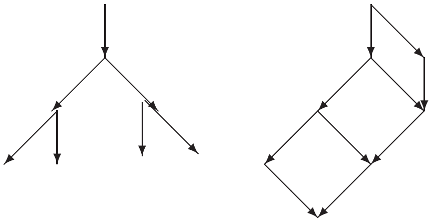


Figure A.15: Directed graphs. On the left, is an acyclic graph, in this case a tree. On the right, a cyclic graph, specifically it is a Directed Acyclic Graph (DAG).

Cliques. Two types of nodes. Bipartate graphs Fig. -A.16

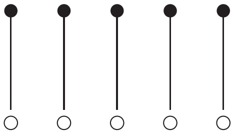


Figure A.16: Bipartate graph.

There are several types of trees such as ordered trees and minimum spanning trees. Trees may also be used to describe sequences of objects (e.g., PAT Trees).

A.3.2. Graph Searching

Many problems such as decision space or a problem space require search. Structured searching in data structures such as binary trees. If there is no index, the graph must be searched by following links and examining nodes. One common trade-off is between breadth-first and depth-first searching (Fig. -A.17).

Tree-searching is useful, for example, in parsing.

Several strategies for searching graphs have been proposed. AI as graph search (3.7.1, -A.7.3).

Heuristic. If value can be assigned at each point, take the one first (Fig. -A.19) but there has to be some criterion for what is the t. Min-max pruning.

The game of tic-tac-toe has a finite number of solutions. Fig. -A.20 shows a game tree. This forms a tree and positive outcomes can be searched. Symmetrical solutions are not shown. While the space for

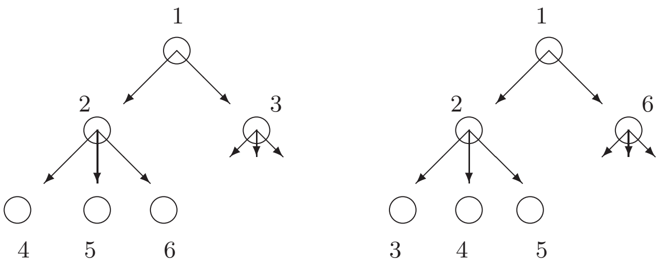


Figure A.17: Breadth-first (left) versus depth-first (right) searching. The numbers indicate the order in which the nodes are searched.

Figure A.18: If values are assigned to each node, those values can be used to guide the search. More complex than simple depth-first and breadth-first search described above, branch-and-bound is t-first searching. As the tree is explored, the likelihood of finding the target under each sub-tree is estimated and the node with the highest value is opened. (redraw) (check permission)

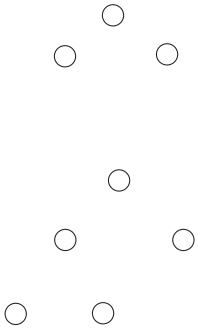


Figure A.19: More complex than simple depth-first and breadth-first search described above, branch-and-bound is t-first searching. As the tree is explored, the likelihood of finding the target under each sub-tree is estimated and the node with the highest value is opened. (check permission)

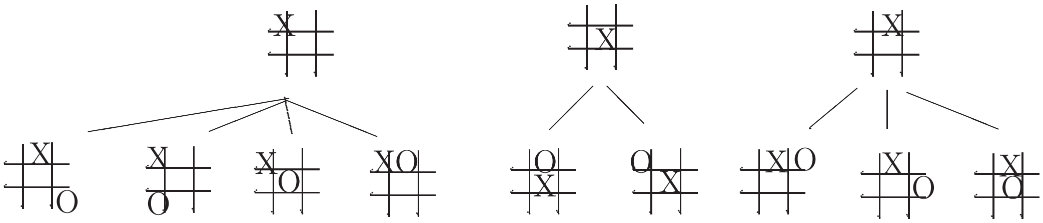


Figure A.20: Fragment for moves in for tic-tac-toe forms a “game space”. Here just the first two moves of a game are shown (other alternatives are symmetrical). We can estimate the value of each move by counting the number of outcomes which lead to winning.

tic-tac-toe is tractable, the game space for chess is far to large to even be calculated. Heuristics might help estimate the value of possible alternatives.

There are many criteria for selecting the value of nodes. Branch-and-bound (-A.3.2). MIN-MAX (-A.9.3). and selection of responses. Insurance.

Searching knowledge structures.

A.3.3. Graph Algorithms

There are many other topics in graph theory that we will consider briefly. General algorithms (-A.5.0). So, for Web characterization (2.6.3) or for validating the coherence of a Web site. Finding the paths through a hypertext. These problems may be viewed abstractly as the connections of nodes among them.

Graph drawing plans the layout of graphs[20]. For instance, when laying out a data map or a flow chart, the graph drawing procedure might attempt to minimize the number of crossings (Fig. -A.21) The goals would be to determine whether two graphs are identical in structure. This includes “graph matching,” “graph homology,” and “graph congruence”.

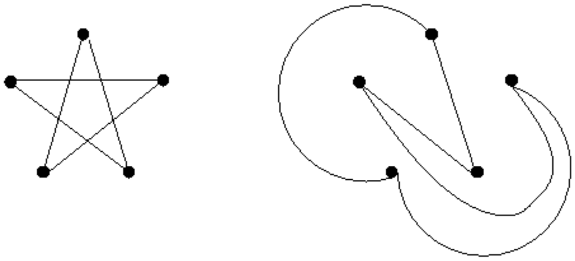


Figure A.21: Two ways of connecting five points. The approach on the left minimizes the length of the lines. While on the right the number of crossings is minimized.

Dynamic graph layout and interfaces. If a display is resized, what is the way to redraw a graph based on it. [45],[52].

One strategy for pruning would cut off those links that do not connect to other links (Fig. -A.22).

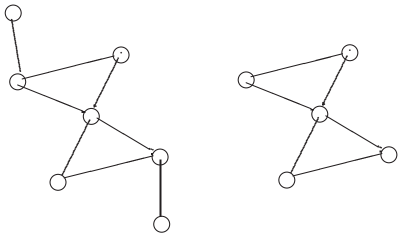


Figure A.22: The graph on the left can be “pruned” to create the one on the right by eliminating those nodes that are connected to only one other node. (check permission)

Beyond pruning to “graph partitioning” (Fig. -A.23). That is, find the place to cut a large graph into two parts. This has been applied to finding Web communities.

Finding the way to connect points in a graph. Spanning trees are trees which connect a set of points. A “minimum spanning tree” is the shortest possible spanning tree (Fig. -A.24).

Pathfinder networks (9.1.3).

A more complex problem is to find optimal paths through set of points. Traveling salesman problem.

A.3.4. Very Large Graphs: From Graphs to Networks

Increasingly, very large scale graphs are being evaluated. Thousands of nodes. A network is a graph in which we consider movement of entities between nodes. Networks may be characterized by some basic properties[8]. One of the most important properties is the distribution of the probabilities of connections between nodes. The simplest has random connection of network nodes. However, some networks have clusters of connected nodes. These are Small-world networks[38] (Fig. -A.25).

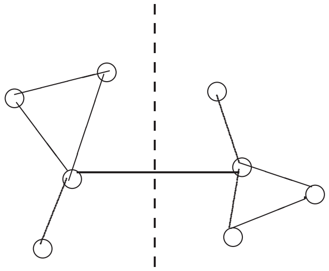


Figure A.23: A large graph may be broken up into small graphs. The partition breaks only one link and leaves roughly equal sub-graphs.

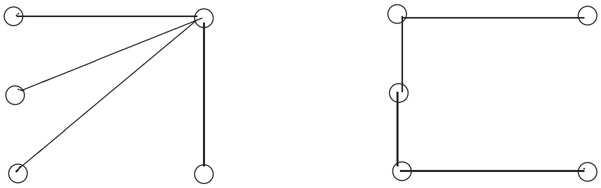


Figure A.24: A spanning tree connects a set of points (left). A minimum spanning tree (right) is the shortest possible tree that connects all the points.

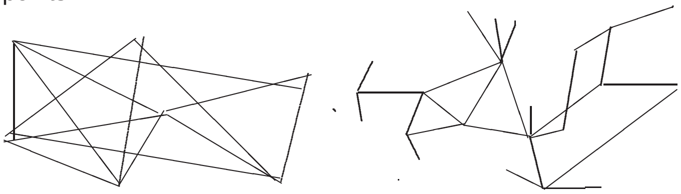


Figure A.25: Graphs differ in the number of short range connections. The graph on the left has random connections. The graph on the right has a preference for connections to some nodes this is typical of many applications such as the Web.

Scale-free structure of network connectivity. Relationship to Zipf's Law.

Graph Complexity

The management of complexity has been a recurrent theme in this text. System complexity (3.8.3). Visual complexity.

Complexity is a challenge. Complexity metrics. Entropy (-A.1.0) as a measure of complexity.

Graph theory (-A.3.0). In Fig. -A.26, the network on the right is clearly more complex than the one on the left. By one common measure, the complexity is X, Y, Z. Thus, complex can be a type of software code metric ((sec:softwaremetrics)). Complexity of software (number of branches and loops). Kolmogorov. Easier to develop and maintain software with less complexity.

A.3.5. Social Network Analysis

We have seen many cases of social information. Purely local interaction. Social networks. Can lead to emergent behavior. Takes a mathematical approach. Who talks to whom (5.1.0).

Characteristics of Social Networks Another type of problem is to determine how close any one item is to any other item in a graph (Fig. -A.28). Because of a classic social psychology experiment done in the 1950's, this is known as the "degree of separation" [50]. In that study, Americans were tested to see how many acquaintances linked people from different regions. This is a similar to the effect of the distance between Web pages (2.6.3). This depends on the "lumpiness" of the graph space. Citation

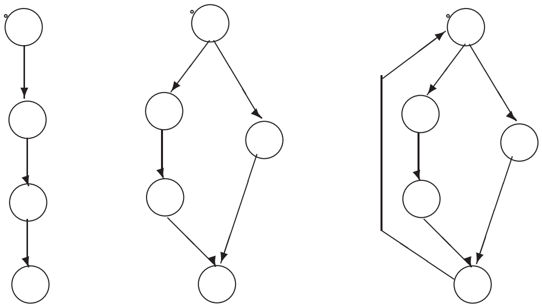


Figure A.26: Branching and looping can be used as a measure of the complexity of a graph. Here, the complexity increases from left to right.

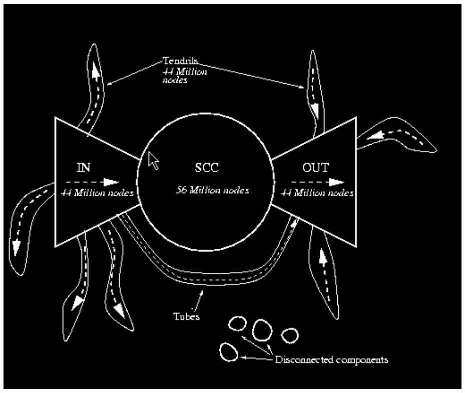


Figure A.27: Massive center for Web interconnections [?]. (redraw) (check permission)

analysis (9.1.2).

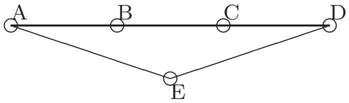


Figure A.28: The degrees of separation count how many steps there are between two points. C is 2 steps from A, but D is also two steps from A when connected by E.

Characteristics of the Social Network Someone at node D is better connected than a person at node J. Characteristics of the individual position in the social network and of the social network as a unit (Fig. -A.30). Consider the communicative patterns of people in the hypothetical communications ("kite") network (Fig. -A.29). In the figure, node "D" has high centrality and node E has high betweenness.

Additional parameters in roles, ease of flow, etc. Correlation coefficient[51].

Epidemics. Inoculation.

Indeed, a critical mass is needed or else, the disease will not be transmitted and will die out.

People on the web can be disambiguated through social networks. Pruning between-ness graph (9.1.3).

Related to PageRank (10.10.2).

Diffusion of Information and Innovation When a new idea or innovation pops up, it gets spread across groups of people. Diffusion of innovations (Fig. -A.32). This is closely related to the social network of

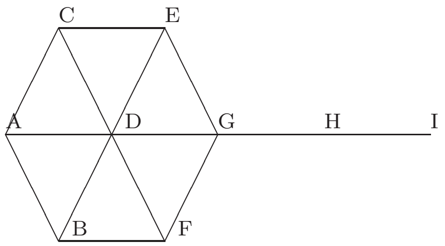


Figure A.29: An idealized communications network — called a “kite” — illustrates the different roles group-members can play in communications. For instance, Person H is essential for I and J to communicate with the rest of the group. (redraw)

Factor	Description
Properties of Individuals	
Centrality	How central is it among other nodes in the network.
Between-ness	Extent to which a node is between two other nodes.
Characteristics of Social Networks	
Density (Coherence)	What proportion of all possible links are actually present?
Cliques	The extent to which subgroups occur.

Figure A.30: Some measures of social networks.

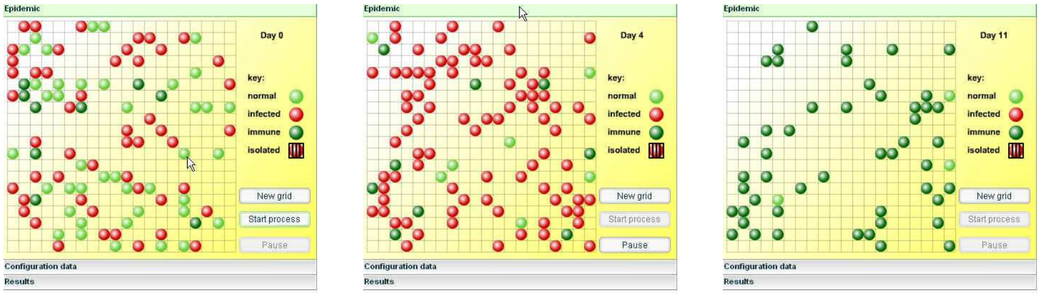


Figure A.31: Epidemic modeling. Normal, infected, and resistant agents are shown. (check permission)(redraw)

the interaction.

Patterns of diffusion. Not simple forwarding. Likelihood of retweeting of political messages. Stickiness and persistence.

In some cases, the forwarding can become a sort of contagion. These are also models for infection and epidemiology including the spread of computer viruses of the spread of human disease. We can talk about contagion and disease vectors. Moreover, these can be blocked with a type of inculcation. Contacts between computers which spread a virus. Following the notion of an epidemic, we can think of a software virus spreading as contagion and we could try to control it with inoculation. In the case of a computer virus, the inoculation might mean applying software patches.

Implemented as an agent-based simulation (9.5.1).

Improvisation as a dynamic model of interaction. Probability of message being accepted. Number of contacts about message. Networking and finding jobs [?]. Probabilistic models. This is too simple a model as the communication and individual action and it must be applied with caution.

A.4. More Models

A.4.1. More Data Models

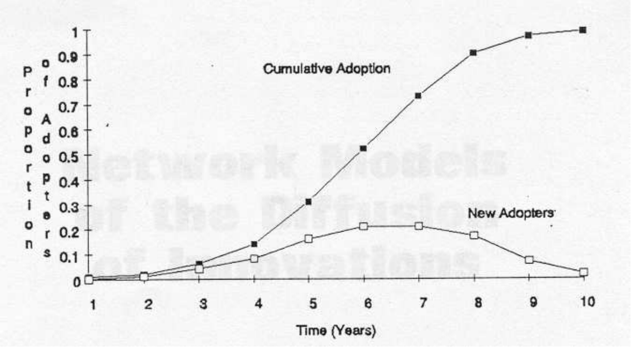


Figure A.32: Idealized curves for diffusion of innovations shows typical curves [?]. (create agent-based simulation). (redraw)

The Relational Data Model: Using Relational Tables to Organize and Merge Attributes

The Relational Model organizes sets of related attributes into tables. Fig. -A.33 shows tables with examples of the entity classes in Fig. 3.55. This use of tables is efficient because it keeps related attributes together. The attributes can be joining entries across tables as needed. Splitting attributes across several tables facilitates efficient storage by minimizing redundancy.

To respond to queries, the attributes often have to be re-combined from different tables. A "key" is an attribute of two or more entities or entity classes that forms a link between entities. In Fig. -A.33, StudioName is a commonality between the two tables; it is an attribute for both entities. Thus, StudioName is a "key," and links the entities STUDIO and VIDEO, and consequently the tables VIDEO and STUDIO. The key guarantees there will be no ambiguity about which rows of the tables to link. The tables are usually optimized with a processes known as normalization. Moving from descriptions of entity classes to specific instances. Attribute value pair: Title="North-by-Northwest"

VIDEO	**Title**	**Director**	**Year**	**StudioName**
	North-by-Northwest	A. Hitchcock	1959	MGM
	Toy Story	J. Lasseter	1995	Disney
	Crouching-Tiger	A. Lee	2002	Columbia

STUDIO	**StudioName**	**Phone**	**Email**
	MGM	800-555-1458	orders@mgm.com
	Disney	800-555-9783	orders@disney.com
	Columbia	800-555-9783	orders@sony.com

Figure A.33: Relational tables and sample values for the VIDEO and STUDIO entities.

Richer Data Models

RDF Data Model Linked data. Often looser structure than formal data models. This can be useful when there are inconsistent systems of metadata.

Temporal Data Models

Modeling Stream Data

A.4.2. Models

System Identification.

A.5. Algorithms

Algorithms describe procedures for accomplishing specific tasks. Algorithmic thinking should be fundamental for education. Algorithms and data structures (3.7.1).

A.5.1. Types of Algorithms

"Algorithms" are procedures for solving problems. Algorithms often need to be coupled with appropriate data structures ((sec:datastructures)). Algorithms have been developed for many of the techniques described in this book. Here we turn to the examination of those abstractions. We briefly discussed many algorithms in the early chapters, and in the more detailed discussions of this chapter we have already considered graph-based algorithms (-A.3.2). Fig. -A.34 shows a table of where some algorithm families are discussed.

Algorithm Family	**Section**
Dynamic programming	(sec:dynamicprogramming)
Encryption	-A.13.0
Graph algorithms	-A.3.2
Compression algorithms	-A.2.0
Machine Learning	-A.11.0
Parsing	-A.5.4
Text and Web Processing	(sec:moretextretrieval)

Figure A.34: Guide to the discussions of some major algorithm families.

There may be several algorithms for completing any given problem, and they may possess different degrees of efficiency in terms of computational cost, memory, or time. Since most interesting problems are complex, it is generally useful to find algorithms that are efficient even when the number of terms gets large. The extra effort required to do addition may increase linearly as the number of terms grows.

Global algorithms take all the data as a unit. These are often the more effective, but they can be very expensive computationally. Other algorithms such as neural networks are "local". That is, the calculations are obtained in steps. A similar dimension is whether the solution is found all at once or whether it is found in iterative steps.

A.5.2. Data Structures

A.5.3. Computational Complexity

For large problems, the complexity can make a big difference in whether a adequate solution can be obtained in the available time. Indeed, we measure the complexity of algorithms in terms of the amount of time they take to complete. Some problems, such as adding a constant to all the members of list are linear. Other problems, such as finding the sum of all pairs of numbers in a list are n^2. The most challenging problems are said to be "NP hard"; their difficulty grows as a polynomial function of their size. Combinatoric explosion.

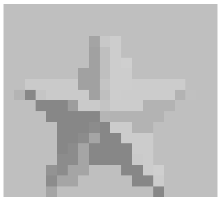


Figure A.35: Comparing algorithm completion time.

A.5.4. Parsing Grammars

Structured objects. Here we explore additional details of the algorithms described earlier (10.4.2, 10.4.2) as well as some other parsing algorithms.

State-Machine Parsing

As noted earlier, natural language can be approximated by a state machine. Extended state machines (3.10.1) can be used for parsing. This works particularly well for formal and simple languages. Specifically, they need to expanded with recursion and otherwise augmented as ATNs.

Fig. -A.37 shows a fragment of a phrase-structure grammar, while Fig. -A.38 shows a very simple lexicon. Specifically, it shows rewrite rules for the sentence "The dog bit the boy". Fig. -A.39 shows the parse tree for this sentence. Collections of tree-structured data — in most cases parse trees — are called a treebank (e.g.,[15]).

Rewrite Rules		**Description**
LHS	**RHS**	
S	NP + VP	Sentences (S) are composed of Noun Phrases (NP) and Verb Phrases (VP)
NP	N, D + N	Noun Phrases (NP) can be composed of a Noun (N) or a Determiner (D) (i.e., 'the') and a Noun (N)
VP	V, V + NP	Verb Phrases (VP) can be composed of a Verb (V) or a Verb and a Verb Phrase (VP)

Figure A.36: Fragment of a phrase-structure grammar. LHS= Left hand side. RHS=right-hand side.

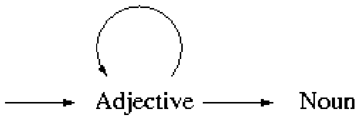


Figure A.37: State machine notation showing that one or more adjectives can be repeated before a noun.

Node	**Lexicon**
Noun	dog, boy
Determiner	the
Verb	bit

Figure A.38: Lexicon for simple phrase-structure grammar example.

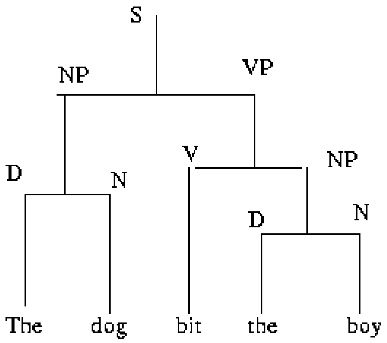


Figure A.39: Parse tree for "the dog bit the boy". (redraw) (check permissions)

"Garden path" sentences. These may require backtracking (3.7.1, -A.7.2).

Other Parsing Methods

Many additional algorithms have been developed.

Chart parsing. Recursive descent parsers. It helps to keep many versions of a parse active.

A.5.5. Hidden Markov Models

Could we formalize the insight we had in Fig. 10.12? Sequential models. Hidden Markov Models provide a statistical technique for modeling sequences. They are weighted automata (10.4.2). Indeed, HMMs may be thought of as a statistical version of grammars. Recall that we used Hidden Markov

Type	Term
S	NP+VP
NP	N, ART+N
VP	V+NP

Type	Term
N	rain, umbrella
ART	the
V	hit

Figure A.40: Simple re-write rules (left) and lexicon (right) for the example grammar.

Step	Description
1.	If the current state can be re-written, use-rewrite rules and increment level.
2.	If state cannot be re-written (at terminal node), check active word against type of the terminal node (active) word.
3.	If that matches, take new terminal word and pop up level and check to see that all tests have been performed there.
3a	If that matches, take new active word and pop up level and check to see that all tests have been performed there.
3b	If there is no match and the state cannot be re-written, back up to previous alternative branches until finding one where a match is possible.
4.	If all the active words have been matched, then the parse succeeds.

Figure A.41: Simple transition-network parsing algorithm.

Step	Active Word	Action/Comments
1	The	Start, expand S to (NP$^{[1]}$+VP$^{[1]}$)
2	The	try NP$^{[1]}$ as (N$^{[2]}$), no match, try next alternative for NP$^{[1]}$
3	The	try NP$^{[1]}$ as (ART$^{[2]}$+N$^{[2]}$), match ART$^{[2]}$, next word, try NP$^{[2]}$
4	rain	try N$^{[2]}$, match, next word, pop up to level 1
5	hit	try VP$^{[1]}$ as (V$^{[2]}$+NP$^{[2]}$), match V$^{[2]}$, next word, check NP$^{[2]}$
6	the	try N$^{[3]}$, no match, try next alternative for NP$^{[2]}$
7	the	try NP$^{[2]}$ as (ART$^{[3]}$+N$^{[3]}$), match ART$^{[3]}$, next word
8	umbrella	try NP$^{[3]}$, match, no more words, pop up to level 0
9	Done	Valid parse!

Figure A.42: Parse of the sentence “The rain hit the umbrella”.

The bear hug created a stir

Figure A.43: Parse for “The bear hug created a stir”. Note that a parser first tries to treat bear as a noun but then has to backtrack and treat it as an adjective.

Models can describe sequences such as the phonemes that represent a spoken word. We have seen many applications of HMMs. HMMs are a type of supervised learning algorithm (-A.11.3) in the sense that the training determines the values of parameters. We have seen applications for parts of speech 6.2.2, 10.4.1 and speech itself 11.3.3.

Selecting an HMM Architecture and Fitting Training Data to that Architecture

The first step is to select an HMM architecture by deciding what constraints can be placed on the HMM. For instance, for speech the models are fed forward. A typical HMM architecture is shown in Fig. -A.44.

The weights for HMMs are usually generated by a supervised learning procedure. Large corpora for training examples. Trying to fit the data into the model. We must have a tagged training corpus. The forward-backward algorithm or the more general, Entropy Minimization (EM) algorithm,[1] is used for training an HMM (-A.1.1) (Fig. -A.45). These combinations may make recognition. These are based

[1] This is also known as the Baum-Welch Algorithm

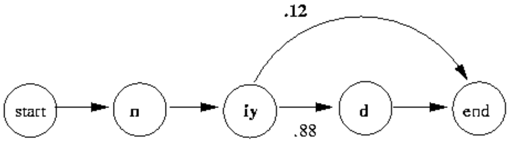


Figure A.44: Repeat of the HMM example which we saw earlier.

on dynamic programming. but they are probabilistic. The continuous stream of speech is difficult to segment into phonemes.

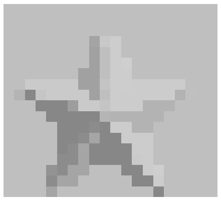


Figure A.45: Forward-backward algorithm for training HMMs.

Matching Sequences to the HMMs

Several HMMs may be developed; then speech samples can be matched to them. This is a kind of model-based recognition with HMM as the model. Specifically, the Viterbi algorithm uses a type of dynamic programming ((sec:dynamicprogramming)) to determine the t-matching sequence (Fig. -A.46). Source-channel model. Information theory (-A.1.0).

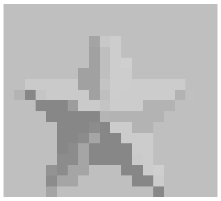


Figure A.46: Viterbi algorithm for matching HMMs.

HMMs are based on Markov models which, generally consider just one previous time step. Although HMMs have proven very successful, more than one time step may need to be considered.

Segment and deal with segments without consideration of the content of those segments. This allows sequential information to be considered. While we might want to use phrases, it may be better to simply use groups of words with a fixed lengths.

A.5.6. Configuration Rules

A.5.7. Optimization and Constraint Processing

Several types of problems.

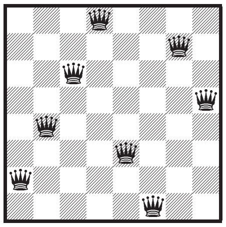


Figure A.47: The 8-queens problem demonstrates the value of algorithms to solve problems that are very difficult to solver by trial-and-error. The queens need to be lined up so that no two are on the same vertical, horizontal or diagonal row.

A.5.8. Version Tracking and Version Management

Keeping track of changes to a document. Fig. -A.49. Dynamic data. Files with periodic updates. Detecting differences in versions. Merge and split. Move. Keeping track of version history.

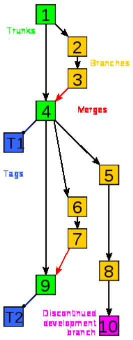


Figure A.48: Versioning. (under construction) (redraw)

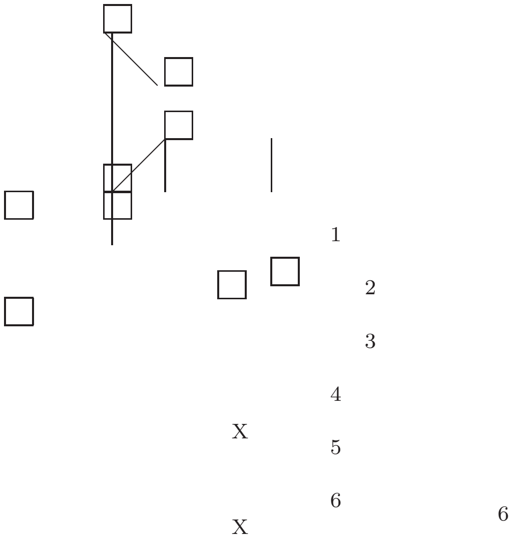


Figure A.49: Keeping track of versions.

Version management for software.

A.6. Additional Search Engine Procedures and Algorithms

A.6.1. Normalization

Preprocessing Text. Inverted indexing. Words (6.2.1). Tokenization, stemming, and normalization. Normalization.

A.6.2. Inverted Indexes

A.6.3. Calculating Term Weights in the Vector Space Model

As described earlier (10.9.2) a text may be represented as a “bag of words” in which the order of the words is not taken into consideration. Compositionality (1.1.3). Fig. -A.50 shows a term-by-document matrix for a hypothetical document collection dealing where each term is just three items.

Here, we use a very simple tf and idf as defined back in (10.9.2). The calculated values are shown in Fig. -A.51.

Similarity and Query Matching

Similarity of documents from word overlaps[75]. There are several ways to measure similarity between

		Document						**Query**
Term		D_1	D_2	D_3	D_4	D_5	*df*	Q_1
1	apple	1	3	2	1	3	5	1
2	banana	4	0	3	0	1	3	1
3	computer	1	4	0	2	5	4	0

Figure A.50: A simplified term-by-document matrix for a hypothetical collection. The number of occurrences of ten terms is shown for six documents. The total number of terms in the document is shown on the bottom line.

		Document				
Term		D_1	D_2	D_3	D_4	D_5
1	apple	1.0	3.0	2.0	1.0	3.0
2	banana	6.7	0.0	5.0	0.0	1.7
3	computer	1.3	5.0	0.0	2.5	6.3

Figure A.51: $tf * idf$ weighting of the documents from Fig. A.50.

two documents or between a document and a query. Some of these approaches simply count the number of overlapping words. Other techniques are based on a calculation of the distance between the documents.

For multi-word queries, a more formal definition of similarity is needed. The "cosine distance" between the query and the documents is calculated separately for each document following Eq. -A.4[2]

$$cosine\ distance\ between\ Document_D\ and\ Query_Q = \frac{\sum_{t=1}^{n}((tf{\cdot}idf)_{t_D} \times (idf)_{t_Q})}{\sqrt{\sum_{t=1}^{n}(tf{\cdot}idf)^2_{t_D}} \times \sqrt{\sum_{t=1}^{n}(idf)^2_{t_Q}}} \tag{-A.4}$$

The match is $Document_3$ (Fig. -A.52). This is reasonable because it has a pattern of $tf{\cdot}idf$ scores that best matches the query idf. They can be improved with more complex tf and df.

Document				
D_1	D_2	D_3	D_4	D_5
0.79	0.36	0.92	0.26	0.46

Figure A.52: The "cosine distance" between the document $tf{\cdot}idf$ values (Fig. ??) and the query idf values. $Document_3$ matches the query t.

$$tf = \frac{log_2(number\ of\ times\ the\ term\ appears\ in\ the\ document + 1)}{total\ number\ of\ terms\ in\ the\ document} = \frac{log_2(t_d + 1)}{T_d} \tag{-A.5}$$

$$idf = log_2(\frac{number\ of\ documents\ in\ the\ collection}{number\ of\ documents\ with\ the\ term}) + 1 = log_2(\frac{D}{D_t}) + 1 \tag{-A.6}$$

The basic $tf{\cdot}idf$ formula includes of the terms in different parts of the document search. However, there are other factors that can also be considered such as query term "prominence". Modern search engines employ other considerations such as term prominence.

[2]. This is derived from the inner product of the Document vector, D, and the Query vector, Q: $cos(\theta) = \frac{D.Q}{|D||Q|}$.

A.6.4. Dimensionality Reduction: Latent Semantic Indexing (LSI)

In many problems there are too many features. Dimensionality reduction reduces the number of features by combining them. The principle of dimensionality reduction has many applications. One useful example of dimensionality reduction is the text retrieval procedure known as "Latent Semantic Indexing" (LSI) which applied dimensionality reduction to the term-by-document matrix of the Vector Space Model (10.9.2). The term "boat" and "yacht" are similar enough that they could be combined. In effect, this creates a statistical thesaurus (2.2.2).

Like the Vector Space Model, LSI usually uses the cosine value for matching. LSI can be used for retrieval[24] and for filtering[29]. As a simple example, in a term-by-document matrix (Fig. -A.53), two clusters of terms may be seen.

	Document					
Term	D_1	D_2	D_3	D_4	D_5	D_6
boat	1	2	0	0	1	0
boats	3	0	7	0	0	0
sailing	4	1	1	0	1	0
water	2	5	3	0	0	0
car	0	1	0	0	6	2
automobile	0	0	0	4	0	5
highway	1	0	0	1	3	0
tires	0	0	0	4	0	2

Figure A.53: In this hypothetical example of a term-by-document matrix, two clusters of documents and terms may be easily identified. One set deals with boats and a second one deals with automobiles. Although the term "boat" does not appear in $Document_2$, the folding of the terms into the reduced-dimensionality LSI space will allow it to be associated with that document.

This procedure should eliminate spurious relationships among the words And focus on the most relevant relationships. As with the Vector Space Model, queries are matched to the documents by taking the cosine distances between the document terms and the query terms. Because this model produces a semantic space, some psychological models have been based on LSI, such as the Latent Semantic Analysis (LSA)[47] of human semantic memory.

Latent semantic indexing uses a linear-algebra technique which is known as "singular-valued decomposition" (SVD). This is related to other statistical techniques such as principle components analysis (PCA) and typically, a high-dimensional space is employed. SVD is also used for eigenfaces ((sec:eigenfaces)).

A.6.5. PageRank Algorithm

The links from one Web page to another provide evidence about similarity of the contents of those pages. Several algorithms have been proposed to demonstrate this (e.g. [**?**, **?**]). However, PageRank focuses only on "authorities".

Here we will consider the details of an algorithm for calculating this. Recall that in Fig. 10.51, we rated pages A and C highly if many other pages point to them. Moreover, if A and C are rated more highly, then B will also be rated highly.

The PageRank algorithm adjusts the rating of pages (nodes) based on the rating of their neighbors with a type of spreading activation (-A.10.3). This is calculated as shown in Eq. -A.7[54]. The Rank of a document, $R(D_i)$, is related to the Rank of all the

documents that are connected to it, $C(N_j)$, where d is a damping factor between 0 and 1. Specifically, the PageRank of P_0 is $R(P_0)$:

$$R(P_0) = (1-d) + d(\frac{R(P_1)}{L_{P_1}}) + ... + d(\frac{R(P_n)}{L_{P_n}}) \qquad (\text{-A.7})$$

Where:
P_0: The target page
P_1,...,P_n are pages linked to P_0
L: outward links from P_0
d: Dampening factor

We can see how this operates for the very simple network in Fig. -A.54 using the values in Fig. -A.55. The overall effect is that activation flows from weakly connected nodes to more highly connected ones.

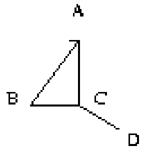


Figure A.54: PageRank calculated for a small network. (use example from MRS)

Node	**Initial Value**	**Ending Value**
A	0.4	x
B	3.1	x
C	1.1	x
D	3.3	x

Figure A.55: PageRank calculations. As would be expected, the activation in the small network accumulates on node A.

A.7. Logic

If we know that "All People Are Mortal" and we know that "Pat is a Person" then it is logical that "Pat is Mortal". Logic is a formal method that supports qualitative reasoning and inference. Logic is used to determine the "truth value" of statements given certain assumptions and inference rules. Like math, logic uses a formal notation and rules. Logic assumes a quantitative (often categorical) processing. Formal logic use ontologies for knowledge representation (2.2.2).

Fig. **??** lists some commonly-used logical symbols.

Types of logic: Description logic. We have seen several examples Knowledge representation. Deontic logic.

Logic is most applicable to discrete categories.

As we will see in (-A.8.1), probability can be used for quantitative inference. Logic versus argumentation (6.3.5).

A.7.1. Symbolic Logic

There are two fundamental types of inference: Deduction and induction. We are generally concerned with deductive inference. This type of logic was originally studied by Aristotle so it is called "Aristotelian logic". Propositional calculus includes deductive statements such as, "If X is true then Y is true".

Categorical Syllogisms

Syllogism is a particular illustration of deduction. For instance, we might attempt to determine the truth of the inferences of a syllogism (Fig. -A.56). Rules of propositional inference.

Statement Type	**Example**
Claim	if students work hard then they are happy
Assertion	all the students in the school work hard
Inference	therefore all the students are happy
Claim	
Assertion	
Inference	

Figure A.56: The first inference is valid (on the assumption that the premises are true) but the second is not.

Truth Functions

We introduced Booleans (3.9.2). Effectively, these are propositions such as "document has term X". We can view Booleans truth tables using the more formalized notation. The XOR relationship is more subtle. The output is TRUE if one or the other input is TRUE, but the output is FALSE if both inputs are TRUE or if both inputs are FALSE. The output is TRUE only if there is discrepancy between the inputs, otherwise the output is FALSE.

XOR		
Input 1	**Input 2**	**Output**
FALSE	FALSE	FALSE
FALSE	TRUE	TRUE
TRUE	FALSE	TRUE
TRUE	TRUE	FALSE

Boolean expressions can be strung together as in the first line of Fig. -A.57. Sometimes, it helps to simplify these expressions into a "normal form". The Conjunctive Normal Form (CNF) .... While the Disjunctive Normal Form (DNF)...

(Chicken AND Dessert) OR (Beef AND Dessert) OR (Chicken AND Coffee) OR (Beef AND Coffee)
(Chicken OR Beef) AND (Desert OR Coffee)

Figure A.57: The expression on these two lines state the same relationship but the second, which is in Conjunctive Normal Form, is more concise.

Sometimes, it is most effective to use a tree to show complex combinations of Boolean relationships. The decision trees we considered earlier were binary OR-trees. They had only OR relationships, but it is also possible to have AND relationships in trees (usually these are indicated with a bar across the choices. Fig. -A.58 shows an AND-OR Tree for the CNF example in the previous table.

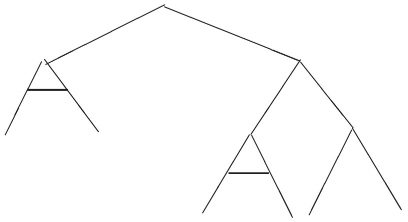


Figure A.58: And-Or trees. The cross-link indicates an AND relationship. (redraw).

Reasoning with Hierarchical Relationships

Inheritance as a model for reasoning (2.1.4).

Problems of multiple inheritance.

Predicate Calculus

Predicates describe the content of propositions. For instance, in the statement $a > b$ the $>$ is the predicate. Thus, predicate logic involves making inferences about statements that include attributes.

Statement	Description	Example
$p \rightarrow q$	If p then q	Apples have seeds.
p	assertion	There is an apple.
q	the conclusion	It has seeds.

Figure A.59: Inference rules.

This often includes the quantifiers "all" and "some". Fig. **??** shows some of the notation. Fig. -A.60 gives an example. If M is a predicate "to be a man," then Mx would be interpreted as x is a *man.*

all students in the school work hard
$\forall x(Zx \rightarrow W)$

Figure A.60: An example of a predicate calculus expression.

Earlier, we introduced ontologies (2.2.2). In the more rigorous sense, ontologies provide the lexicon of the predicate calculus.

Assertion links.

Frames as a generalization of hierarchies.

Frames

Frames are a way of representing entity classes. However, unlike Entity Classes from the ER model, they usually apply to general world knowledge. Still, the frame-slots are a lot like attributes. Fig. -A.62

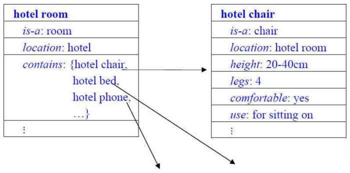


Figure A.61: Frames. (redraw)

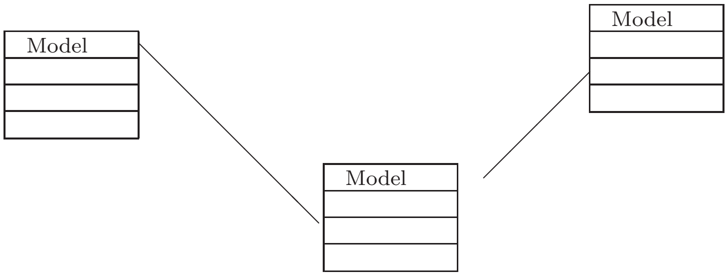


Figure A.62: Frames. (under construction)

Thus far, we described logic based on inferences involving evaluating the validity of statements about specific instances. This is "first-order predicate calculus". Second-order predicate calculus examines the validity of statements about relationships. Representing facts.

A.7.2. Complex Logical Inference

Implications such as double negative elimination.

Inferences create new propositions. Given a set of statements, we can try to "chain" them together to draw inferences. The objective is generally to find a path from the initial state to the final goal. We could start with the premises and build inferences toward a goal (forward chaining).

Or, we could start with the goal and attempt to reason backward (backward chaining). Forward chaining expands all propositions in order to find all possible implications. Backward chaining identifies the goals and then progressively decomposes those goals into sub-goals Fig. 3.40. This process is similar to "means-ends analysis" (3.7.1). These can also be viewed as examples of bottom-up processing and top-down processing (10.1.5).

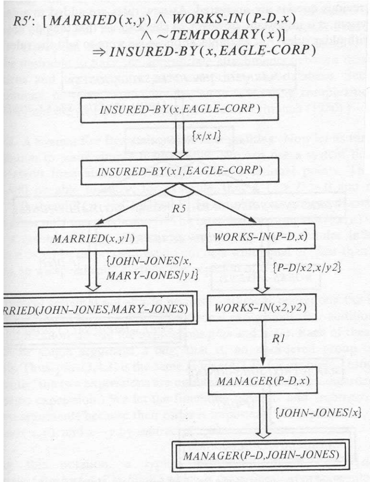


Figure A.63: Backward chaining. (check permission)

A.7.3. Knowledge Representation and Logic Programming Languages

Knowledge representation (2.0.0). Formal languages for knowledge representation and logical inference.

Declarative Logic Programming Languages

Several logic programming languages have been developed. Fig. -A.64 shows some examples of Prolog statements about kinship. Given these definitions and assertions, we could answer questions such as "Is there a child whose parent is Eve?"

Statement	**Explanation**
woman(eve)	Declare there is a woman named Eve.
man(adam)	Declare there is a man named Adam.
child(abel)	Declare there is a child named Abel.
mother(M,C):-woman(M), parent(M,C)	Define that a mother is a woman who is a parent.
father(F,C):-man(F), parent(F,C)	Define that a father is a man who is a parent.
mother(eve, abel)	Declare that Eve is the Mother of Abel.
father(adam, abel)	Declare that Adam is the Father of Abel.

Figure A.64: Prolog is a computer programming language design to perform logic operations.

Procedural Models: Production Systems

Approaches such as Prolog are "declarative" These may be distinguished from "procedural" models (Fig. **??**). Declarative specifies rules: this includes logic. Procedural is a specification for what is legal such as production systems.

Production systems are based on Condition-Action pairs. That is, if certain conditions are met, then the production "fires" and the action is executed. Productions may also be thought of as sets of IF-THEN statements. SOAR[46] is another production system language ((sec:productionsystem)) that allows decomposition of goals and selection of rules. Fig. -A.65 traces the steps of a SOAR program as it is executing. Potentially, the priority of SOAR rules can be "learned" by storing those productions for later use that were most effective. SOAR allows machine learning (-A.11.0). by chunking (4.3.5).

```
0: ==>G: G1
1:    P: P1 (farmer)
2:    S: S1
3:    ==>G: G3 (operator tie)
4:      P: P2 (selection)
5:      S: S2
6:      O: O8 (evaluate-object O1 (move-alone))
7:      ==>G: G4 (operator no-change)
8:       P: P1 (farmer)
9:       S: S3
10:        O: C2 (move-alone)
```

Figure A.65: SOAR Problem Space Computational Model trace[46]. There are Goals(G), Proposition(P), States(S), and Objects(O). (check permission)

Expert Systems

Expert systems use inference and reasoning for practical applications. These are often 'rule-based, that is they are based on logical inference. For instance, they may such as production systems to inference. Unfortunately, expert systems tend to be "brittle". That is, they may work well for the situation for which they were developed, but do not generalize well to new situations. These can also be decision support systems (3.4.2) decision support systems but there is a danger of inappropriate inference. Furthermore, they may exceed the application domain of the system. For instance, the Aegis attack (Fig. -A.66).

Technology failures ((sec:techfailures)).

Figure A.66: Aegis.

Fuzzy logic and probability of belonging to a set (Fig. -A.67).

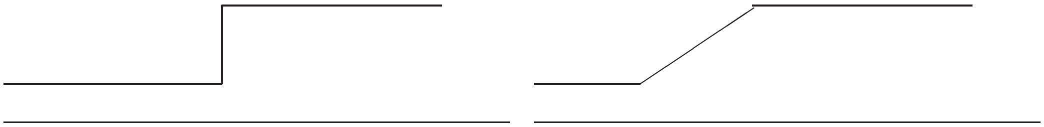


Figure A.67: Normal local (left) and fuzzy logic (right).

A.7.4. Representation and Reasoning with Beliefs

Earlier, we discussed beliefs as an aspect of social psychology (4.5.0). In one of the senses of formal description of beliefs.

In a model with several agents, those agents may have models the world, of each other, and other agents views of the world. (Fig. -A.68). A person may believe something that is not true. Or, it may simply be impossible to verify. If I believe the world is flat...
I believe in dragons....

Belief is different from confidence.

Propositional attitudes[19].

Logic and beliefs[28]. Reasoning about uncertainty[36].

John knows his name. John believes that today is Tuesday.

Belief vs. belief systems.

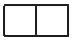


Figure A.68: Beliefs.

Confidence in results.

A.8. Probabilities and Probabilistic Inference

Alternatives to logical inference which was discussed in the previous section. Toward plausible reasoning. Induction. Uncertain results. Sampling. Hypothesis testing.

A.8.1. Basic Probabilities

While logical inference is based on symbolic inference; it is also possible to make a probabilistic inferences. Let us briefly review probability. Eq. -A.9 shows the product of two probabilities. This essentially an AND operation. For instance, this probability of getting a 2♣ AND 3♠ in one draw is 1/13 + 1/13. Eq. -A.9 shows the sum of probabilities. This is essentially an AND. For instance, the probability of getting a K♣ AND Q♣ in successive draws is 1/13 * 1/13 (assuming you the cards are replaced after each draw). Note that the symbol ∩ is the same as *and*.

$$P(A\ and\ B) = P(A \cap B) == P(A) * P(B) \tag{-A.8}$$

$$P(A\ or\ B) = P(A) + P(B) \tag{-A.9}$$

We can also define conditional probability which is, for instance, the chance of “Event A given Event B” that can be written as $P(A|B)$.

$$P(A|B) = \frac{probability\ of\ both\ Event\ A\ and\ Event\ B}{probability\ of\ Event\ B} = \frac{P(A \cap B)}{P(B)} \tag{-A.10}$$

A.8.2. Bayesian Prediction

Learning conditional probabilities. This is a common technique for machine learning (-A.11.0). It can also be seen as a type of knowledge representation.

Bayes Rule

If we have expectations about how attributes predict membership in a category we may also be able to determine that likelihood that objects in the category will show possess those attributes. Specially, if we know P(A—B), P(A), and P(B), we can determine P(B—A). This is known as Bayes Rule and it is the basis of learning about the features relevant for doing categorization.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{-A.11}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{-A.12}$$

$$P(A \cap B) = P(A|B)P(A) = P(B|A)P(B) \tag{-A.13}$$

Suppose there is a 0.5 probability documents with the word “training” in them also have the word “education” and the probability of the word “education” occurring in the documents is 0.8. If we know the word “education” is in a document, what is the chance that the word “training” is in that same document? Eq. -A.14:

$$P(\text{“}training\text{”}|\text{”}education'') = \frac{P(\text{“}training'' \cap \text{“}education'')}{P(\text{“}education'')} = \frac{0.5}{0.8} = 0.625 \qquad (\text{-A.14})$$

Bayesian Classification

Bayes Rule can also be applied to categorization. Thus, if we know the frequency of a set of categories and we know the frequency with which terms occur in documents belonging to those categories, then we determine the probability of a new document belonging to a category given the terms it includes.

If we know that an object has a certain attribute value, we might ask “what is the probability that object or event belongs to a given category?” This can be determined with an extension of Bayes Rule (Eq. -A.15). For instance, Eq. -A.15 shows the probability of belonging to Category 1 (C_1) given that Attribute 1 (A_1) has Value 1 (V_1).

$$P(C_1|A_1V_1) = \frac{P(C_1 \cap A_1V_1)}{P(A_1V_1)} \qquad (\text{-A.15})$$

This may be generalized to multi-attribute categories[26] (Eq. -A.16).

$$P(C|A_1V_1, A_2V_2, ..., A_NV_N) = \frac{P(C_1 \cap A_1V_1, A_2V_2, ..., A_NV_N)}{P(A_1V_1, A_2V_2, ..., A_NV_N)} \qquad (\text{-A.16})$$

Bayesian Networks

Attribute-based conditional probabilities.

Updating causal networks[57].

Used in text retrieval Fig. -A.69

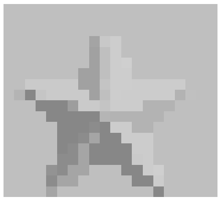


Figure A.69: Bayes Network visualization.

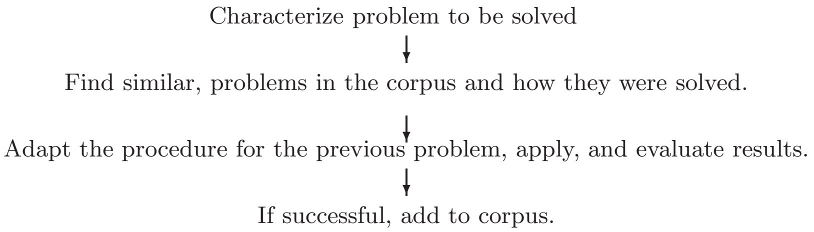
Information gain in Bayesian calculations.

A.8.3. Case-Based Reasoning (CBR)

Case-based reasoning (CBR) attempts to find relevant examples to generalize from rather than trying to develop a comprehensive statistical model[17]. For instance, when modeling the path of hurricane, it may be more useful to examine previous similar hurricanes rather than trying to rely on complex simulations. The researcher must still find effective features and representations (Fig. -A.70). Retrieve, re-use, revise, return. Sets of examples may be maintained in case libraries.

Formal Descriptions of Cases

Setting, Actor, Goals, Sequence. Case-based reasoning (-A.8.3).



Characterize problem to be solved

Find similar, problems in the corpus and how they were solved.

Adapt the procedure for the previous problem, apply, and evaluate results.

If successful, add to corpus.

Figure A.70: Path of case-based reasoning.

A.9. Formal Models for Decision Making

Choice and decision making have appeared earlier (3.4.1).

Logic, Inference, Planning, and Learning.

The contents of representations are sometimes presented directly to the user. In many other cases, they must be reassembled. Doing things with representations.

Algorithms for both recognition and generation.

The simplest task is making binary YES/NO decisions. For instance, we could detect the possibility of a terror attack from a cumulative set of data.

A.9.1. Signal Processing

A signal carries information in the information theory sense (-A.1.0). A signal can be lost if there is too much noise. As an example, think about trying to listen to radio station when there's static. You need to concentrate to detect the music. The simplest approach to determining whether a signal is present or absent. Fig. -A.71 shows distributions of signal and noise. The signal-to-noise ratio determines the ease with which the signal can be detected. Consider trying to hear a telephone ring in another room of your house. It is much easier to detect the telephone when there is not any background noise such as the radio playing or the shower running.

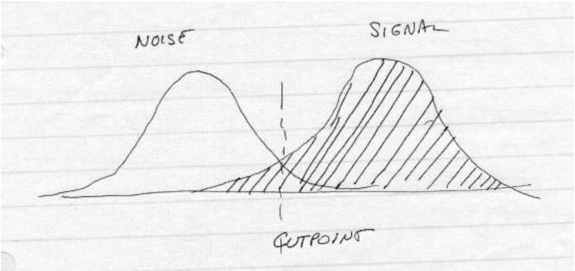


Figure A.71: Distribution with overlap and a decision threshold. If the threshold is moved to the left, more signal events can be detected but more errors are also made. However, it the threshold is moved to the right, fewer errors are made but more signals are missed. (to be rendered)

A.9.2. Signal Detection

Detection is simply a decision whether a signal is present or absent. If the signal and noise are similar, it may be difficult to tell them apart. Success in monitoring the occurrence of events (signals).

A measure of the success of detecting signal is developed as follows. Fig. -A.71 also shows signal and noise distributions. Also shown is a cut-point, which is the threshold at which an observer would decide the signal (i.e., the telephone ring) was present or absent. There are four possible combinations of signal and user responses (Fig. -A.72). The cut-point is normally selected to minimize the number of errors, but other strategies for placing the cut-point could also be considered.

		Actual Signal	
		Present	**Absent**
Observer's Judgment about Signal	**Yes**	Hit	False Alarm (False Positive)
	No	Miss (False Negative)	Correct Rejection

Figure A.72: 2x2 table for signal detection. The observations might not be accurate since they might be due to noise, as suggested by Fig. A.71.

This is a type of classification problem.

It is harder to understand somebody when they are talking in a noisy environment than in a quiet place. The level of the signal compared to the amount of noise is known as the "signal-to-noise ratio". Two factors determine the signal-to-noise ratio: The difference between signal and noise and the observer. This statistic is known as d'.

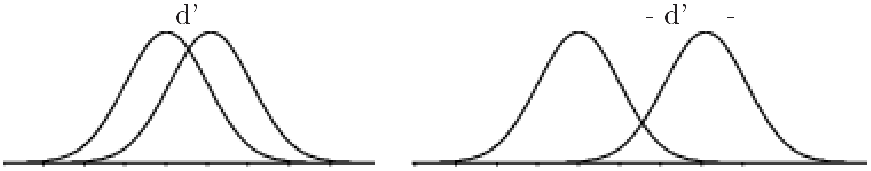


Figure A.73: Sometimes the noise is similar to the signal (left) and sometimes it is clearly different (right). When it is similar, it takes a very sensitive detection device to accurately separate the noise from the signal. (label distributions)

We should keep the ratio of False Positives and False Negatives constant. We can do characterize observers as to whether they have a bias toward false positives or false negatives. Compare the two distributions and determine how good is an observer at telling the difference. Response operator characteristic (ROC) curves (Fig. -A.74).

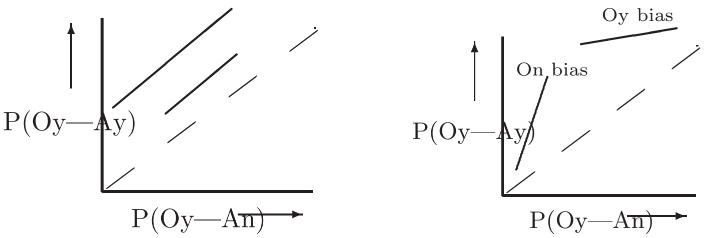


Figure A.74: The Response-Operator-Characteristic (ROC) diagram shows how an observer responds given varying probabilities of Yes and No responses and the signal is varied. The diagonal represents chance performance. As shown in the ROC diagram on the left, the further the ROC curve is from the diagonal, the better the discrimination. The diagram on the right shows an analysis of whether the operator has a bias toward responding "present" a bias toward responding "absent".

E-measure for information retrieval.

Generally an issue for recognition processes. Signal detection is closely related to categorization. It determines whether an object belongs to a given group or not[35]. The properties of signal detection.

Recognizing Category Membership

As we have seen, categories are widely used in information systems (3.9.1, 4.3.0). We briefly discussed choice strategies earlier (3.4.1).

A.9.3. More Game Theory

A well-known collective action game is the Prisoner's Dilemma, which is illustrated by the values in Fig. -A.75. Imagine two prisoners A and B, who were partners in the same crime but who are being interrogated separately by the police. If they both confess, they may get a moderate punishment (-3), but if one prisoner confesses while the other does not the one who confesses will get a light punishment (-1), and the one who does not will get a heavy punishment (-5). However if neither talks, there is no direct evidence and they might both go free (+5).

		Prisoner A	
		A does not talk	A talks
Prisoner B	B does not talk	5/5	-1/-5
	B talks	-5/-1	-3/-3

Figure A.75: In the "prisoner's dilemma," the payoffs for each prisoner depend on the behavior of the other prisoner. The cells of the table shows payoffs to each of the two prisoners.

Game theory can also be used to explain long-term interaction[65]. During the Cold War, the theory of Mutually Assured Destruction (MAD) developed based on game theory. the claim was that the only stable equilibrium was the point at which each side could destroy the other. Collective action games seek to analyze the decisions made by individuals when the outcomes of those decisions are affected by the decisions of other individuals. In many cases, one person or the other will have a clear advantage. However, the players and will tend to stabilize at an equilibrium point that has advantages for both players. The Nash equilibrium is the solution for which the players would not change their strategies even knowing the choice of the opponent.

		Country A	
		No Bomb	Use Bomb
Country B	No Bomb	0,0	-1000,10
	Use Bomb	10,-1000	$-\infty$,$-\infty$

Figure A.76: Game theory table for Mutually Assured Destruction (MAD). (revise)

Strategies in risky situations. The most directly applicable competitive strategy for making decisions involving other individuals and/or imperfect information is one that picks outcomes that maximize the benefits and minimize the risks. This is known as a "min-max" strategy. A person who has to make a choice among a number of approaches may analyze the chances of favorable and unfavorable outcomes. This is reasonable since it assumes that the opponent will also attempt to maximize his/her benefit.

Max-min as a fairness strategy.

		Payoff			
		A	B	C	D
Possible Outcomes	Gain	+3	+5	+6	+6
	Loss	-4	-4	-5	-6

Figure A.77: According to a min-max strategy, the options with the minimum loss are selected and from those, the options with the maximum gain are selected. Thus, option B would be selected. This has the minimum loss for that gain and the highest possible gain. However using a max-min strategy, C would be selected. This has the minimum possible loss and the maximum possible gain.

A.9.4. Subjective Multi-Attribute Utility

People may have their own utility for different attributes and we should include these subjective utilities in our models the choices those people will make. Getting reliable values of subjective utility is not easy[27]. This is an example of "scaling" (e.g., [?]).

$$value\ of\ object_i = \sum_{j=0}^{j=N} (attributeValue_{ij} * utility_j) \tag{-A.17}$$

We would like to estimate utility when multiple attributes are involved. If we know preferences, we can work backward and estimate the utility. Multi-attribute decision theory is a mathematical means of analyzing decisions in which there are several competing variables to consider. In multi-attribute decision theory (or multi-attribute utility), each variable is assigned a particular utility value according to its importance and they are all plugged into a mathematical formula to determine what combination of variables produces the most desirable outcome. For instance, a person might chose between two models of cars based on their attributes (see Fig. -A.78).

	Type of Car		
Dimension	Compact	Sports Car	Sedan
Price	3	1	2
Fun	1	3	1
Safety	2	1	3

Figure A.78: Several attributes of cars could be assigned values based on their favorability. A score of "1" is low on that dimension and a score of "3" is high.

	Type of Buyer	
Dimension	**Yuppie**	**Family**
Price	1	2
Fun	3	1
Safety	2	3

Figure A.79: Subjective utilities for two types of buyers. Higher numbers mean that the dimension is more important for that type of buyer.

	Yuppie			**Family**		
Dimension	Compact	Sports Car	Sedan	Compact	Sports Car	Sedan
Price	1*3	1*1	2*1	3*2	1*2	2*2
Fun	1*3	3*3	1*3	1*1	3*1	1*1
Safety	2*2	1*2	3*2	2*3	1*3	3*3
Overall preference	10	12	11	13	8	14

Figure A.80: The Yuppie buyer will prefer the sports car while the Family buyer will prefer the sedan.

A.9.5. Voting Systems and Elections

Voting involves the allocation of units decision units across candidates and rules for combining those units (8.4.3). A voting system needs to accurately reflect the voters' preference. Perhaps surprisingly, that does not always happen with simple majority rules ting. Fig. -A.81 shows one example of a complications introduced in three-way race. To solve this problem, a variety of voting criteria have been developed (Fig. -A.82). These may allow multiple votes per individual and preference rankings of several candidates[62]. End-2-End (E2E) electronic voting security. Open source voting software.

Elections

System of voting and related procedures for determining government officials. Non-partisan supervision of elections.

Candidate	**Strength of Preferences**			
	$Voter_1$	$Voter_2$	$Voter_3$	Mean
A	0.40	0.05	0.35	0.27
B	0.35	0.30	0.30	0.32
C	0.15	0.45	0.25	0.28

Figure A.81: Majority rule may not result in the most preferred (on average) candidate being elected. If each voter is allowed to cast only one vote, Candidate A would be elected as the top choice of the majority of the voters. However if voters cast votes in proportion to their preferences, Candidate B would win.

Type	**Description or Example**
Majority	One vote per voter. Winner needs more than 50%
Plurality	One vote per voter. Winner is the candidate with the highest number of votes.
Borda	Voters rank order the alternatives. Candidate with the highest average rank wins.
Approval	Cast one votes for each candidate the voter would accept. The winner is the candidates with the highest number of votes.
Cumulative	Each voter has multiple votes. These can be cast all for one candidate, or spread across candidates. The winner is the candidates with the highest number of votes.
Instant run-off	Successive run-offs narrow the field of candidates.

Figure A.82: Several types of policies and criteria for elections.

A.10. Mathematical Models

We have discussed many types of models and mentioned mathematical equations. Discrete math versus continuous models versus continuous models. Relationship between mathematical models and scientific models (9.2.3). Thus, an important distinction is between linear and non-linear models. These differ in the power of the representation (1.1.2). Machine learning, clustering and neural networks (-A.11.0). Models in science (9.2.2). Deterministic versus probabilistic models. To an extent, all models can be thought of as mathematical functions. Levels of models. Ordinal, Interval, For instance decision trees are qualitative models. These are representations based mathematical functions. Free parameters. Over-fitting. Fitting data: Model + error.

A.10.1. Linear Models

If we believe that some simple linear process accounts for an effect, we might attempt to fit the data for that to determine the parameters. the linear model from the some data. An effective approach is often to find the line that is the least-squares distance (Fig. -A.83). Underlying linear model plus error in measurement.

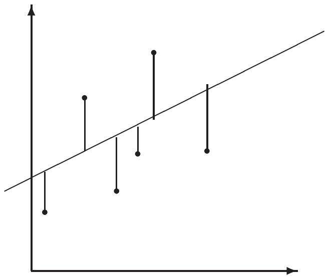


Figure A.83: A straight line is fitted to a data points. A common approach is fitting with “least squares” which finds the line that minimizes the sum of the square of the distance from the data points.

Even simple models can provide immense analytical help. Fig. -A.84 illustrates how simple algebraic models can be used to determine what is the combination of production capacities to produce two separate products. The two left panels describe the production capacity of two types of cars (product 1 and product 2). The right panel then uses linear algebra to resolve the constraints posed by factors of production.

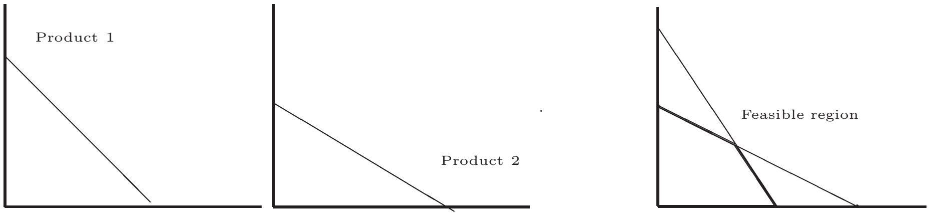


Figure A.84: Linear models can specify constraints and simple combinations of these constraints can be calculated. On the right, the two linear constraints are intersected. If the first line shows the maximum level of X and the second line shows the maximum level of Y, then the combination of the two shows the "feasible region" BELOW the intersecting lines. In other words, that area within which parametric tradeoffs are possible.

A.10.2. Non-Linear and Dynamical Models

In some cases, the interaction between the components is often very unpredictable. When two adaptive systems interact, they form a "dynamical system". These are also called "co-evolutionary" or "mutually-causative" systems. The evolution is determined by the direction the pair of systems take. An example would be a person interacting with another person — the actions of each affect the other. While some of these systems are chaotic, others are stable.

Sometimes linear equations are good representations for a process; sometimes a more complex, nonlinear equation works better. We have already seen non-linear models used for mathematical simulations (9.5.4). While linear systems are very powerful, many systems are non-linear. The representations are mathematical equations. Linear modeling, which assumes that all effects can be modeled with straight lines, is effective only to a point. Equations with exponents. These models are the foundations of complex systems.

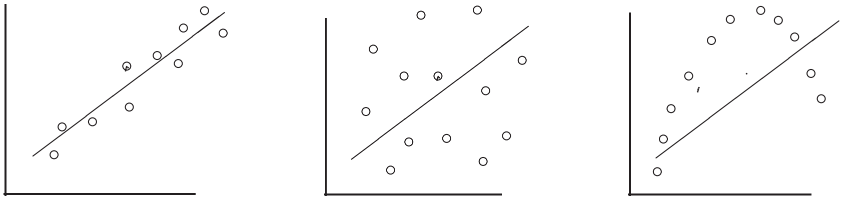


Figure A.85: A straight line can be an effective representation to describe a set of points (left). However, the line is less satisfactory if the points are scattered (center) or if a curved line may be a better representation (right). A representation that allows curved lines will also be more complex.

Power Laws

These are a family of common non-linear functions. For instance, the long-tail (8.12.5) follows a power law.

$$y = x^z \tag{-A.18}$$

Some typical power law curves are illustrated in Fig. -A.86

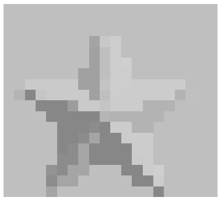


Figure A.86: Power laws.

80-20 rule.

More about the long tail (8.12.5).

Long-Tail Distributions

Praeto

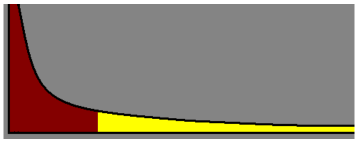


Figure A.87: Long tail. (redraw)

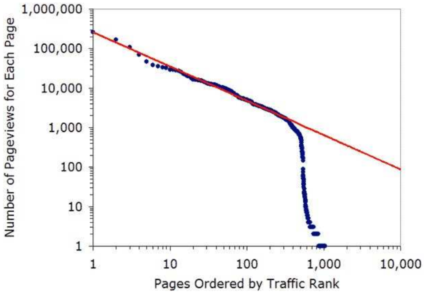


Figure A.88: When we expect data to fit a power law, deviations from the predicted pattern may indicate underlying problems. Here, a drooping tail in eCommerce data suggests that the may not be enough representation of low-frequency items. (check permission)

Zipf's Law Closely related is a simple mathematical function known as Zipf's Law (Eq. -A.19) gives accurate descriptions of word frequencies. Zipf's Law states that the frequency of observations for a word of a given rank number P_r, is equal to a constant, k, divided by the rank, r:

$$P_r = \frac{k}{r} \tag{-A.19}$$

Application of Zipf's Law. Example of word frequency. Fig. **??**.

1.	the
2.	
3.	

Figure A.89: Example data for Zipf's Law.

Fractals and Self-Similarity

Chaotic systems have no stable solutions. However, some do exhibit a property known as "self-similarity". Self-similarity suggests that there is a repetition of a pattern across several different scales. Fig. -A.90 gives two examples of this property. The self-similarity in some of these patterns generates complex patterns. There are applications of fractals in image generation and compression.

Simulation of non-linear and complex systems. Simulated annealing.

Set point with a comparator.

There are systems with the factors are interlocking. Simple feedback with a controller. Control theory (Fig. **??**). Such models are too simple. Unlike adaptive models in which the representation itself changes.

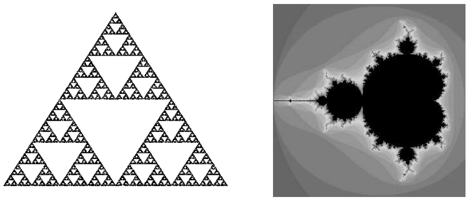


Figure A.90: Self-similarity is illustrated on the left in the Sierpinski triangle in which equilateral triangles are carved out of larger equilateral triangles[32]. (check permission) On the right is a fragment of the Mandelbrot Set which shows a more complex self-similarity. Zooming in shows essentially identical patterns that are repeated at the finer levels of granularity.

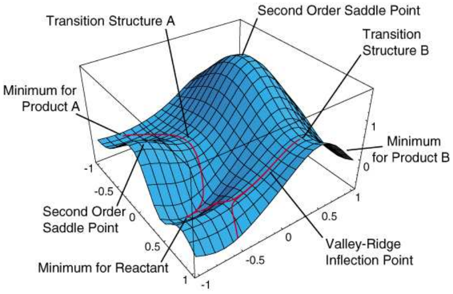


Figure A.91: Energy surface. Finding an energy minimum. Simulated annealing.

Sometimes we need several interacting equations to model a system, These form "dynamical systems" Sometimes these equations converge to a solution and a system with feedback will maintain homeostasis around a control point. In other feedback systems diverge and no solution is possible. Those that converge reach a single "fixed-point" equilibrium are said to be attractors (Fig. -A.92). Sometimes the equations do not converge to a single point but follow a regular pattern across several solution points. In a few cases, there is no simple pattern to the solution.

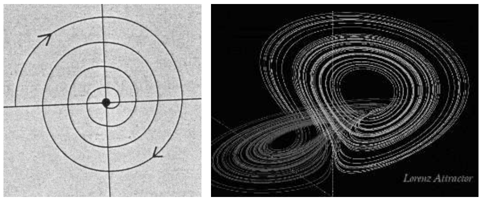


Figure A.92: Trajectories of an attractor (left) and a strange attractor (right)[32]. (check permission)

There are sometimes complex systems. Chaos comes from large differences due to small changes in initial conditions. An example is the "butterfly effect" in which an apparently insignificant event in one part of the system can be amplified to have a major impact later in the system's evolution.

Critical phenomena. Emergent phenomena.

Punctuated equilibrium.

Sometimes simulations are used to model these systems but one has to be careful about the accuracy of the simulation.

Hysteresis.

Dynamical Systems

Non-linear systems with feedback. These are sometimes called "co-evolutionary systems". System dynamics (-A.10.2). The two components interact together and their combination reaches a unique state.

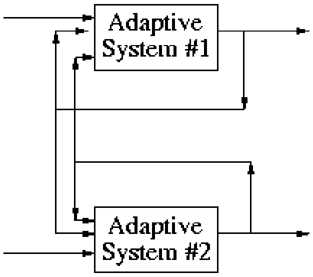


Figure A.93: A pair of interacting adaptive systems can be treated as a single complex system. The output of one system forms part of the input for the other system. (re-orient horizontally)

Some equations model how a system changes through time. Some of these consistently converge to a single point across many trials; other systems diverge. It is possible to describe complex interactions with sets of equations. These are relatively easy to solve when the functions are all linear, However, the solutions are more complicated when the equations are non-linear.

System Dynamics

System Dynamics include feedback but also "stocks". For instance, in a supply chain analysis (8.12.1) the goal might be to keep an inventory of parts roughly constant while they were being used in manufacturing. Or, as illustrated in Fig. A.94, the level of population could be modeled as it is increased by births and decreased by deaths. Moreover, there is a positive feedback such that the more people there are, the more births there will be. On the other hand, the more deaths there are, the fewer deaths would be expected in the future.

The interlocking feedback loops often make change extremely difficult.

This is like a data flow diagram (3.10.1). Such models can provide insight into why some processes are so resistant to change[66].

There is no exact computational solution for these models. Numerical analysis.

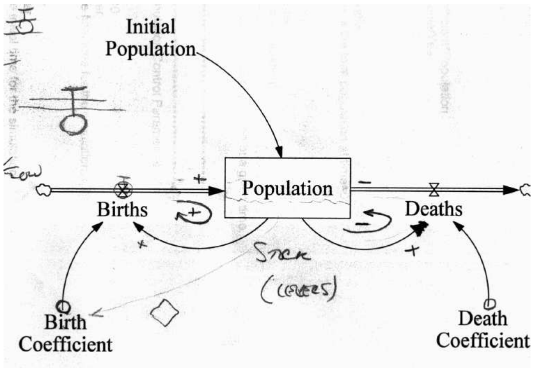


Figure A.94: Flow in a population-growth diagram. As the population increases, both births and deaths will increase. (redraw)

Exampless of supply chain applications (8.12.1).

Causal Models

System Dynamics Models An important class of models are those which can represent rates of change, in other words, for models which are highly non-linear. For instance, we might like to model how population size changes as food supply changes. These have feedback. These are more difficult to model.

Because of the interaction of factor approximation must be done by numerical analysis. System-dynamic models (-A.10.2). Simulations (9.5.0). Causal loop diagrams (Fig. -A.95). Complex systems (-A.10.2).

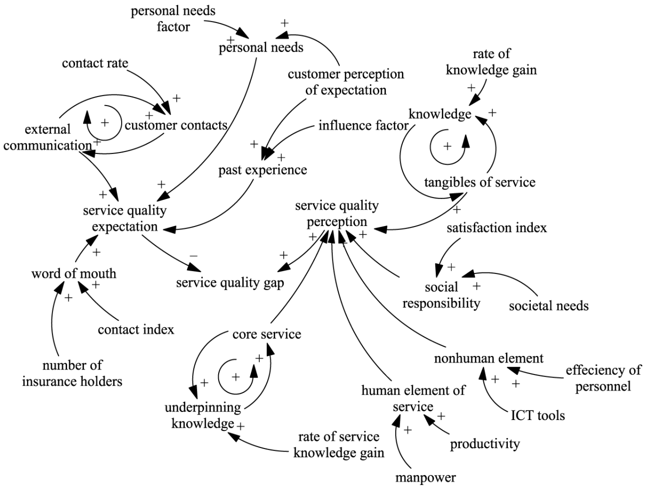


Figure A.95: Qualitative causal loop model. (redraw)(check permission)

Causation is intergral with explanation (6.3.4)especially explanation in science ((sec:sciexplanation)). Implications for social science (4.4.2). Bayes models for causation.

Structural Equation Models Causation (4.4.2) can be inferred based on a model. For instance, in Fig. -A.96 Compare to DAGs and Bayesian Networks.

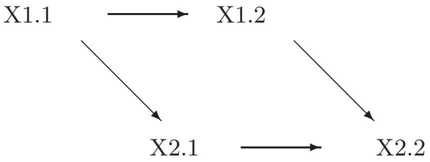


Figure A.96: Structural equation models can help to validate assertions made about causal relationships.

Determining latent variables[6].

A.10.3. Network Flows and Related Problems

Flow in a Network

One application of graphs (-A.3.0) is to examine flow through the network. Queuing theory. Calculating costs of routing. Traffic on city streets (Fig. -A.97). Predicting congestion. Volume and ease of flow. Dynamic models for optimizing flow. This also has implications for Social Network Analysis.

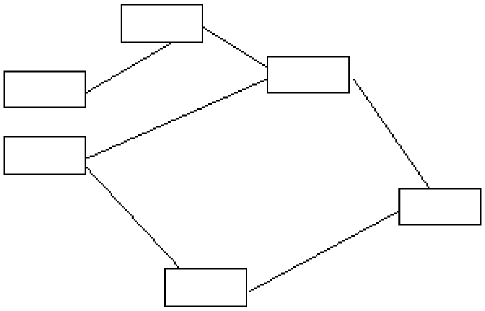


Figure A.97: Network flow. (redraw by hand)

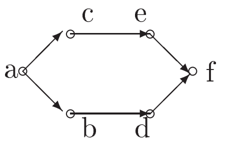
A related problem is based on the arrival time by cars at a toll booth and how long they have to wait. Queuing theory.

Internet routing tables. Packets in a network.

Spreading Activation

Because so many systems are modeled as graphs, we can explore the spread of activation. The nervous system can be thought of a network in which neurons (nodes) are connected by links. When a person thinks about one concept, related concepts often seem to come to mind. For instance, if I think about my dog, I might also think about the park near my house where I walk the dog.

This priming effect could be modeled with activation which spreads across the graph. Suppose there are six nodes connected as in Fig. -A.98. An impulse starting from neuron a would go to both b and c. In turn, the impulse would be transmitted from those two nodes on to nodes d and e and then finally to node f. Suppose further that only 70% of the activation gets through with each hop so that 0.49 * 0.70 = 0.24 and then 0.24 + 0.24 = 0.48. Many additional parameters could be applied to this model such as only on/off neurons, a transfer function (including a maximum activation), and speed of decay of the activation.



	time step		
neuron	1	2	3
a	1.00	1.00	1.00
b	0.00	0.70	0.70
c	0.00	0.70	0.70
d	0.00	0.49	0.49
e	0.00	0.49	0.49
f	0.00	0.00	0.48

Figure A.98: The spread of activation from neuron a to neuron f across three time steps.

A.10.4. Agent-based Models

Using independent agents with local rules to obtain a stable solution in a complex system. This is a natural extension of social networks (5.1.0).

One strategy for this uses cellular automata. Some simulations is best done with connected "cells". We call simulations with these "cellular automata". Computer models used for weather forecasting are extremely complex.

Agent-based simulations.

One example of a cellular automata is the Game of Life[30]. (Fig. -A.99).

1.	If a cell is dead and if three of its neighbors are alive then it comes alive.
2.	If a cell is alive and are two or three are alive, it stays alive.
3.	Otherwise, a cell dies or remains dead.

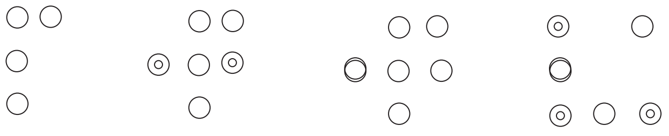


Figure A.99: Rules for the Game of Life (top) and an example of its use. The filled circles are newly born.

Artificial life models are artificial systems that behave in a fashion similar to the organisms. Random mutation and natural selection are elegant means by which individual organisms, species, and ecosystems interact to produce structured change at all levels and, typically, increasing complexity. The field of artificial life uses these same principles to design "environments" in which programs interact to produce change in themselves and the environment itself. This is an example of cybernetic evolutionary modeling.

Computer viruses as a form of artificial life.

Axelrod

Cellular automata can be used to take the modeling of living organisms more literally, "artificial life". Analysis of biological processes (Fig. -A.100).

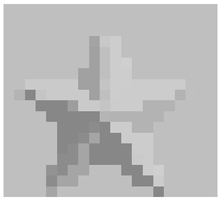


Figure A.100: An example of artificial life. (check permission)

Swarm Intelligence.

Self-organizing systems.

A.11. Learning Mechanisms and Machine Learning

We have already explored learning in several places. For instance, we have considered human learning (4.3.5) and other adaptive systems. Cognition and learning (4.3.5). By "machine learning" we mean algorithmic learning. Simple categorization is sometimes considered learning; however, here the focus is on learning in which an entirely new representation is developed. In most cases, the machine learning algorithm is trained in one phase and its performance is tested in a second phase. Generally need large datasets for statistical approaches to linguistics. Generalization. Over-learning.

A.11.1. Learning Mechanisms

We have learning processes in many places. Learning as taking advice. Coaching.

We briefly described human learning earlier (4.3.5); we can look more closely at learning. According to a behaviorist definition, human learning can only be demonstrated as a change in behavior since we can never be sure what representations people use.

Types of learning can be based on the conditions in which they occur. "Learning by doing" or "Learning by observation" Another strategy for discussing learning is to focus on changes in representation. unsupervised and supervised. There are many possible applications such as grammar induction or learning how to recognizing speech acts. Furthermore, machine learning can be applied to adaptive interfaces.

For human learning, we cannot know in detail how human learning occurs by inspection develop a model for it (4.3.5). However, we may program a computer to do simple learning.

Reinforcement learning. Some tasks such as learning language seem to involve feedback. Learning language from positive examples.

Conditioning. Loud noises and bright lights have a direct physiological impact as an "unconditioned stimulus". Other stimuli may be conditioned by association with the UCS.

Reflection and consolidation seem to be important for human learning (5.11.2).

Here, we will focus on unsupervised and supervised learning.

In some cases, a short-cut can be learned. Chunking. Learning patterns of checkers[63].

Game space (-A.3.2). Skipping a deep search in a game tree. Parameter learning.

A.11.2. Unsupervised Machine Learning

The classification processes discussed earlier assumed a predefined category system. Unsupervised learning systems attempt to "discover" the structure of the underlying similarity of a collection of objects. For instance, sets of abstracts for documents might be identified. We might think of this as creating plausible categories.

Agglomerative clustering versus partitioning approaches.

Classifier.

Emergent concept learning (1.1.4).

Quantitative and Hierarchical Clustering

Hierarchies are particularly effective for organizing information. Cluster analyses tries to find a hierarchy to fit data. A graphical presentation of the output of a hierarchical cluster analysis called a "dendrogram" (Fig. -A.101).

From clustering to classification.

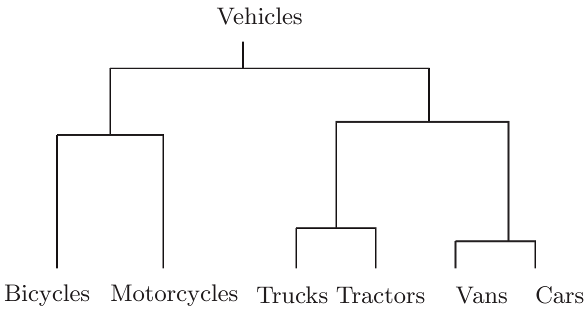


Figure A.101: Dendrogram that might be obtained from a hierarchical cluster analysis on the distance between six types of vehicles. Ideally, the clustering will end up with cleanly separated categories.

Qualitative Clustering and Decision Trees

While the most common type of clustering is qualitative, other clustering techniques have been proposed which are based on quantitative attributes. Decision trees were introduced earlier (3.4.1). Simple decision trees can be created by hand, but more complicated ones are better made with specialized tools. Two of the better-known approaches for developing decision trees are Classification and Regression Tree Methodology (CART) [22] and ID3 [58]. This proceeds in merging from the bottom up (Fig. -A.102). These methods work well for data sets that are linearly separable but models such as back-propagation (-A.11.4) are better for problems where non-linear partitions are possible.

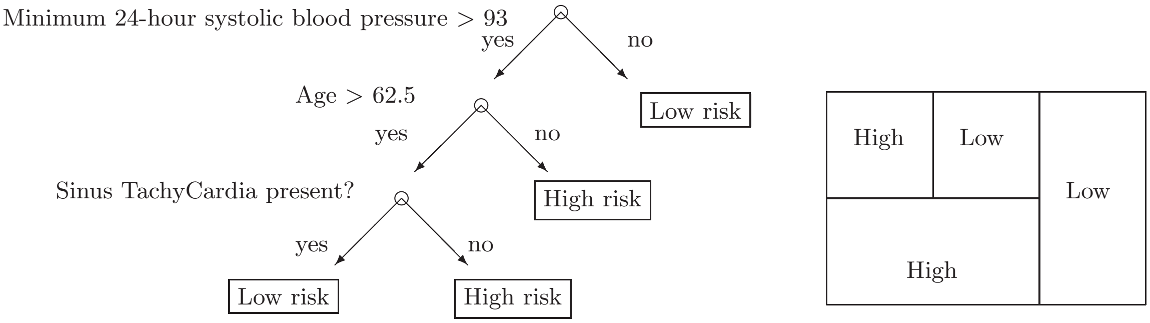


Figure A.102: CART decision tree for[21] and a schematic of the partitions it makes for treatment of hospital patients.

Scaling

Measurement (9.3.0)/ Metric and non-metric measurement.

Transformations.

Transitivity.

Multidimensional Scaling (MDS). Multi-Dimensional Scaling (MDS) is related to cluster analysis. MDS attempts to find the fit of data of high dimensionality into a quantitative method which can be used to form non-hierarchical clusters a lower-dimensional space.

Self-Organizing Systems and Maps

When a crystal forms, the atoms or molecules in it align themselves in highly-ordered patterns. This is a type of self-organizing system. Typically, these have local units that organize into larger, more coherent patterns. Society, the Web, and life itself are all generally considered to be self-organizing systems. This is a type of unsupervised learning.

FMRI evidence for concepts separate from language [**?**].

A.11.3. Supervised Machine Learning: Learning Category Membership and Similarity

Supervised learning algorithms use feedback about the results of an action from the environment to improve performance. Because classification is so ubiquitous, we often think of learning as improving the quality of classification. Supervised learning requires a representation to be updated so that the next time the behavior is emitted, it is done better.

This is sometimes called learning by trial and error. From design to requirements.

Active learning. A process of improving categorization. For instance, we might select the optimal training set.

Feedback can be either positive or negative. Instrumental learning is learning which helps a person to accomplish some goal.

The procedures also differ in their representation. This section focuses on neural networks, but other well-known supervised learning procedures include genetic algorithms (-A.11.6), Hidden Markov Models (-A.5.5), and Bayesian learning (-A.8.2).

For instance, text categorization (10.6.1) might use Bayesian techniques.

Issues for Supervised Learning. How much training? How good is generalization. Transfer (4.3.5).

Classifiers.

Supervised algorithms generally require several training cycles. By gradually improving the model, the algorithm may be able to perform better on later tasks. This is a process known as “hill climbing”. Not every task is amenable to every supervised learning algorithm. For instance, if the process of gradually improving the weights reaches a local minimum which the algorithm cannot pass to reach the global minimum.

Supervised learning algorithms use feedback. Some algorithms will not necessarily converge and show an improvement. Measures of learning include generalization to new situations. Another problem is over-generalization, which is learning about the details of a specific training set and missing effective generalization.

Training strategies. Training the network. Successive approximations and learning.

When there is a complex model, This is known as “credit assignment”. Sometimes, it may be difficult to determine which factor contributes to the result. Therefore, it will be difficult to update the right part. Changing the model.

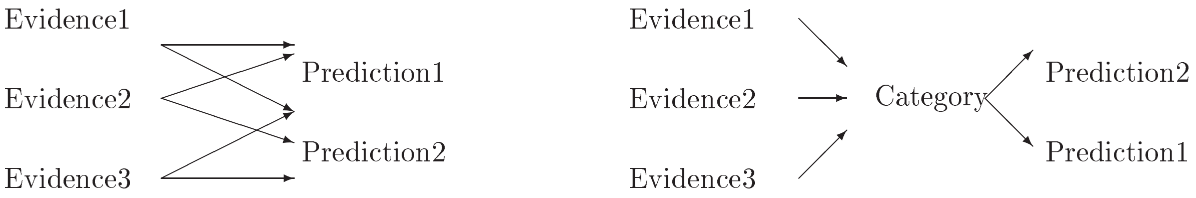


Figure A.103: For some types of predictions, it is helpful to categorize the input values but in other cases, using numerical values without categorization is more effective.

A.11.4. Learning in Neural Networks

Neural networks may be thought of as a style of computation. Some approaches to learning are modeled loosely on biological systems that show learning.

Neural networks are computational models which can be applied to learning algorithm in which the computation and representation are distributed across interconnected nodes. Neural networks are loosely modeled on neurons in the nervous system. Each neuron is a very simple processing unit. Typically, each neuron is active and makes a contribution. Thus, the computation is parallel and emergent. Neural networks are numeric and do not explicitly model symbols and concepts. Thus, they provide an alternative to symbolic processing models. It remains unclear whether these system can learn to manipulate symbols.

Neural nets and pattern recognition. Modeling the responses of conversational agents. Neural networks are used in many ways including classification. This can be used in general data mining (9.6.5).

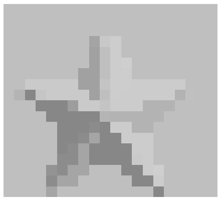


Figure A.104: Distributed representations. (redraw)

Learning Different Types of Representations

Overview of Neural Networks

The basic neural network model is composed of neurons connected by activation pathways. Each neuron combines the inputs from the activation paths and applies a transfer function to determine how much activation will be presented. The neural networks can learn representations that characterize the patterns of inputs they have received.

Most retrieval systems employ indirect indexing terms to point to the content. An alternative is to have the content serve as its own index. Content-addressable memories.

“Neurons that fire together wire together.” Typically, information is represented in the neural networks by the weights and activation algorithms. The representations developed by neural networks are distributed and difficult to examine. They are an excellent example of non-symbolic processing. However, neural networks have been criticized for not being able to yield explanations for how they reach decisions. More often, supervised learning algorithms gradually change the weights. Neural networks support associative learning (4.3.5). A simple assumption, which is known as Hebbian neurons, states that when two neurons are both active at the same time (i.e., associated), then the strength of the link between them is increased[37] (Fig. -A.105).

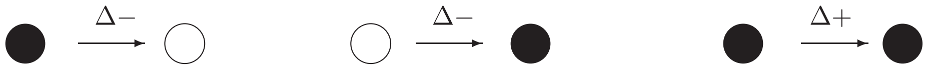


Figure A.105: According to the Hebbian learning model the weight, or strength of association, between two neurons should be increased (Δ+). when the neurons are reacting the same way. That is, when they are both ON (filled circles) (left). If they are reacting differently (center and right), the strength of the association between them is decremented (Δ−).

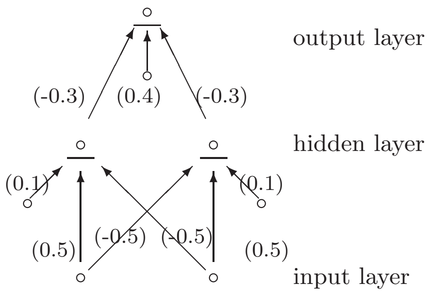
Back-propagation Algorithm

To demonstrate even simple reasoning, a learning system should be able to at least learn basic Boolean operations. The Boolean XOR is similar to the Boolean operations described earlier (3.9.2). The XOR was originally believed not to be learnable by neural networks. The back-propagation algorithm[61] became particularly well-known when it was demonstrated that it could learn the XOR logic function (Fig. 3.58). This is a type of non-linear regression.

There are many ways collections of neurons may be connected. This is the foundation for the representation. Fig. -A.106 shows a simple three-layer neural network. The input values are shown by the weights of the links, which connect them to the hidden-layer neurons. This is, essentially, a bottom-up process (10.1.5). The three-layer model is particularly effective for data reduction in which the number of hidden units is small compared to the number of inputs or outputs.

The basic idea is that the weights are updated so the network is more likely to produce the desired result after the update. Each training trial has two phases. The forward-propagation follows a type of spreading activation network (-A.10.3). However, the basic spreading activation approach is adapted with the inclusion of bias units, negative weights, and synapses with transfer functions (Fig. -A.106).

The level of activation on the hidden layer is determined by a simple formula which integrates the activation from the inputs. The same process is repeated starting with the hidden units to obtain the activation on the output. For a network which has already been trained, the output values should match the targets.



		Pattern			
layer	neurons	activation	activation	activation	activation
input	left	0.0	1.0	0.0	1.0
	right	0.0	0.0	1.0	0.0
hidden	left	(0.1) 0.0	(0.6) 0.8	(-0.4) 0.2	(0.1) 0.0
	right	(0.1) 0.0	(-0.4) 0.2	(0.6) 0.8	(0.1) 0.0
output		(0.4) 0.2	(0.8) 0.9	(0.8) 0.9	(0.4) 0.2
target		0.0	1.0	1.0	0.0

Figure A.106: Forward-propagation in three-layer neural network. Note that the weights (shown in parentheses in the schematic) are preset to value which solve the XOR. The activation spreads from the input layer through the hidden (middle) layer to the output layer. Activations are collected at synapses, which are shown by horizontal lines. A transformation is applied to the synapse activation shown in parentheses in the table to produce the neuron activation.

For the neural network to demonstrated learning (i.e., for the weights to be updated) we can use the difference from the target along with the strength of the activation on each weight by a small amount. These corrections are made on the weights from the outputs back to hidden units and then on the weights back to the input units.

A.11.5. Deep Learning

Feature extraction.

A.11.6. Genetic Algorithms

DNA is the representation for adaptive biological systems. In the biological systems, learning is accomplished from mutation followed by natural selection. Since we know that biological species adapt through evolution, it may be possible to imitate them. This process can be simulated with binary strings representing a gene pool (Fig. -A.107). Changes are introduced by mutation of the binary string. "Cross-overs" are a type of mutation in which segments of two strings are swapped (Fig. -A.108). Natural selection can then be simulated by selecting those mutated segments that provide better responses to the problem the initial patterns.

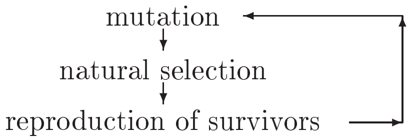


Figure A.107: Steps in evolution are emulated by genetic algorithms.

Initial Patterns	**Ending Patterns**
1 1 | 0 0 0	1 1 | 0 1 1
1 0 | 0 1 1	1 0 | 0 0 0

Figure A.108: In genetic algorithms, new bit patterns may evolve by a process of "mutation" and "natural selection". An example of crossover from a genetic algorithm is shown; the last three bits have been flipped.

A.12. Biological Basis of Human and Social Information Processing

A.12.1. Biological Bases of Social Behavior

Social brain. Aggression. Empathy.

Animal models for social behavior. Fig. ??.

Figure A.109: Chimp grooming. (check permission)

A.12.2. Brain Science

Why brain science is relevant for information science.

While we have generally focused on the use of information rather than the underlying infrastructure. For human information processing we have considered cognition (4.3.0) but not the brain. Here, we investigate that. Neurology. Cognition systems. Hierarchical sensory processing. Metabolic cost for cognition.

Scenario visualization [?].

Plasticity. Language learning up to a certain age. Sensory and brain development.

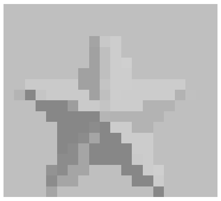


Figure A.110: Basic brain structures.

Cognitive metabolic costs. Metabolic costs in multitasking. Motor behavior (4.2.4).

Perhaps some overlap of brain inputs helps to give people the distinctive capabilities[64]. Face blindness.

Spatial Brain

Spatial brain. Grid cells for location.

Brain Structure

Modularity of brain structures.

Left-handedness.

synapses > neurons > network > maps > nervous systems

Macro Structure Regions for vision, emotions, motor control.

There is some cross-talk among neurons in the brain. Priming. Even apparently cross-talk between structures. Holding a hot cup of coffee affects rating of the warmth of other people.

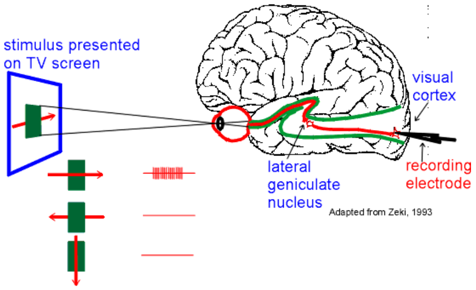
The human brain is vastly different from silicon computers and their programming. The brain is a mass of neurons which are inter-connected by an even larger number of axons.

The physical structure of the brain shows a lot of specialization. Right brain versus left brain [?]. identifying brain function of different brain hemispheres. Left brain tends to be logical and the right brain tends to be intuitive.

Hippocampus. Spatial neurons.

Social brain. Face recognition. Empathy.

Fear and aggression. Emotion from the amygdale.



Adapted from Zeki, 1993

Figure A.111: Hubel and Wisel neurons. (check permission)

Mirror neurons and empathy.

Motor behavior (4.2.4) and sensation (Fig. -A.112).

Micro-structures. Mirror neurons.

Brain science and language learning. Broca and Wernicke's Areas are important in language.

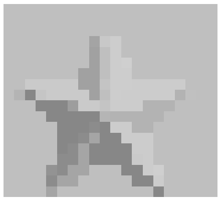


Figure A.112: Brain motor behavior.

Neurons Neurons.

Synapses are adaptive. Neurotransmitters. Dopamine.

The ability of the human brain to develop new representations seems to change with growth.

Brain Function

Furthermore, the functions of many regions of the brain can be clearly identified. Right brain versus left brain[11]. The left hemisphere of the brain is generally associated with speech. One part, Broca's Area, is involved in speech and language production. While another part, Wernicke's Area, is involved in speech understanding. Moreover, these may be related to language difficulties such as dyslexia (4.9.3).

Unreliable components (i.e., neurons) produce generally coherent thinking.

Visual features and visual search.

Magnetic resonance imaging (MRI) fMRI which measures increased blood flow for different cognitive activities.

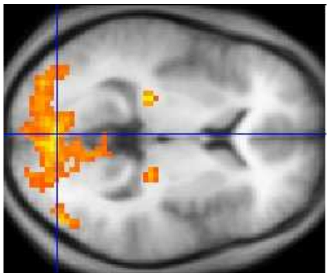


Figure A.113: Functional magnetic resonance imaging (FMRI) has proven very useful for determining which parts of the brain are most involved in high level cognitive processing. (check permission).

Some of these studies have revealed specialized structures of the brain. Regions of the visual cortex specialized for faces, places, bodies[41] (Fig. -A.114).

Hippocampus and episodic memory.

Expert chess players versus novices show activity in different regions of the brain when playing chess.

Neural plasticity.

Mental imagery and vision.

Pain.

Music[48].

Category-specific cells. Grandmother cells. Grammar cells. Cells which respond to stimuli which have been attended to a lot.

Consciousness and intentional behavior.

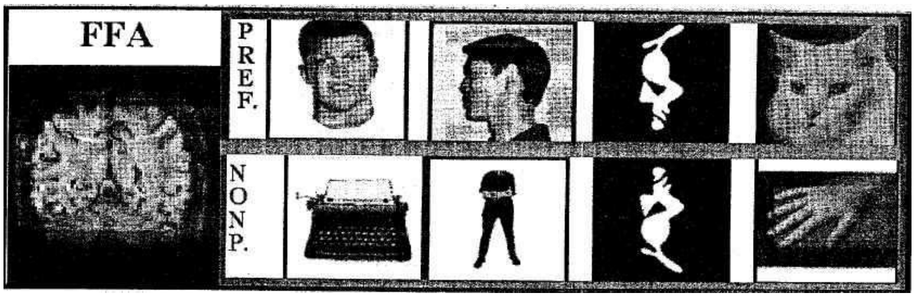


Figure A.114: Face detection cells vs frequent item cells.[41]. (check permission)

Neural network models (-A.11.4) and broader modeling of neural circuits.

Modular system with lots of feedback[70].

Sleep and memory consolidation[7].

Multiple memories. Amnesia difficulties of forming long-term memories.

A.12.3. Affect and Emotion

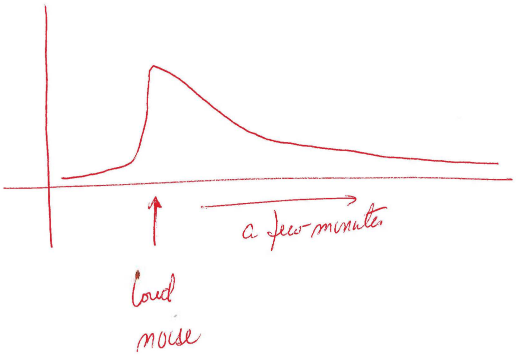


Figure A.115: Typical time-course for physiological arousal. A loud noise may cause an "fight or flight" reaction". (redraw)

Oxytocin.

Addiction

Multiple competing forces. Opponent process model of addiction. Pleasure and stress.

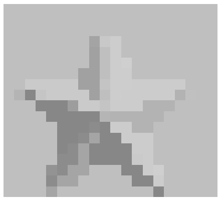


Figure A.116: Model for addiction.

A.12.4. Learning

As with the complexity of neurons themselves, there are many mechanisms for learning. There is both plasticity and wired-in learning. Some studies show that neural organization of information continues as late as 18 years of age.

A.12.5. Brain Simulation

Machine learning (-A.11.0). .

Neural Prosthetics Neural prosthetics.

Neuro-technologies.

Moral judgment,

A.13. Encryption and Cryptography

Encryption is a foundation for information security and applications such as privacy (8.3.1) and ecommerce (8.12.5). We briefly introduced encryption earlier (-A.13.1); here we extend that. These algorithms are triggered by a "key" which is a large number that sets the algorithm. Many of the most robust encryption protocols are based on the difficulty of factoring combinations of prime numbers. Amount of comutation is a consideration for routine use. Brute force attack to break encryption.

Figure A.117: Bletchley house: The site of British code breaking work during World War II. (check permission)

Hidden Markov models (-A.5.5) can be useful for code-breaking. Specifically, they can help to identify non-random processes.

A.13.1. Encryption

Encryption is a base technology which can facilitate security. Encryption scrambles data, making it difficult to intercept and read. Encryption supports for information assurance. When important data can be easily and illegitimately copied, and other information can be as easily forged, it is natural to look for technological solutions to the problems raised by such activities. For all practical purposes, modern encryption algorithms cannot be broken. In a sort of arms race, sophisticated technologies for protecting information often produce sophisticated attacks by people seeking to breach those safeguards. The encryption algorithms may be embedded in a service to aid in information security.

Secret Codes

Codes and encryption protect information from being seen by people who do not know the key. Some of the simplest codes are "substitution codes" in which one letter is replaced by other letters. The code shown in Fig. -A.118 is formed by shifting each letter 13 positions in the alphabet. These rotated letters are substituted for the original letters. We might easily guess that the rotated letters v, b, and a are among the most common in the language since they appear twice in the coded word. In fact, we see that these letters represent i, o, and n. With sufficient samples of text, such simple codes are easily broken.

e	t	a	o	i	n	s	r	h	l	d	c	u	m	f	p	g	w	y	b	v	k	x	j	q	z

Figure A.118: Rank order of the letters in the English (Latin) alphabet based on their frequency.

original	i	n	f	o	r	m	a	t	i	o	n
ROT13 coded	v	a	s	b	e	z	n	g	v	b	a

Figure A.119: The letters of the word in the first line are shifted by 13 letter positions (ROT13) in the second line.

Single-Key (Symmetric) Encryption

Modern encryption schemes are much more difficult to break. One family of encryptions is symmetric; that is, the same algorithm key can encrypt and decrypt these codes. The Data Encryption Standards (DES) (-A.13.2) is called "symmetric" because the encryption and decryption keys are the same (Fig. -A.120). Without knowing the key, the only practical way to break these codes is by testing all possible values for the keys. Whether, and how quickly, the algorithm can be broken depends on the size of the factors and the speed of the computers trying to break it.

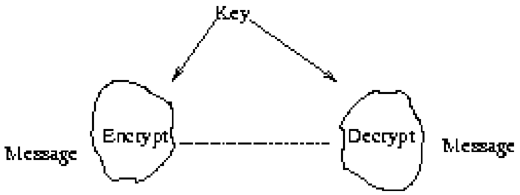


Figure A.120: Symmetric-key encryption uses the same key for encryption and decryption.

Public-Key (Asymmetric) Encryption and the Public Key Infrastructure (PKI)

The public-key algorithm uses two asymmetric keys. One of the keys encrypts messages while the second decrypts them. The details of the algorithm are given in -A.13.3. Most often, the public key algorithm is used to prove that information is from an authentic source. It could be a digital signature or a stamp to validate a Web site. In this type of application, the encryption key is kept secret and the decryption key is made freely available. If a Web site is able to be read using the decryption key they we can be confident it was encrypted by the holder of the private key. It is also possible to publish the encryption key and keep the decryption key secret. In this latter approach, anyone can encrypt a message and send it to the holder of the decryption key, but only that person can read the message. Signing certificates.

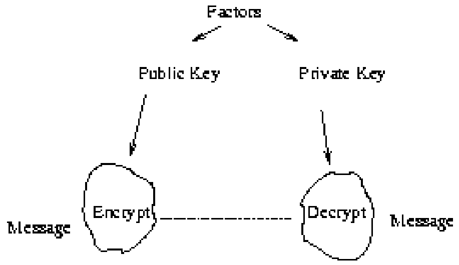
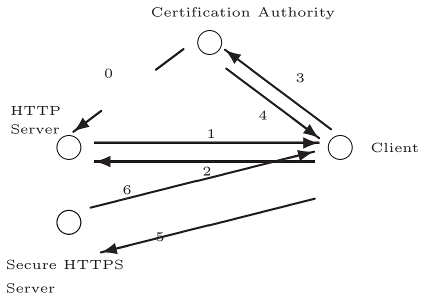


Figure A.121: Public-key encryption is asymmetric with one key to encrypt (write) the file and a second key to read it.

Beyond encryption algorithms, an infrastructure is needed to allow distributed computers to exchange information securely. A certification authority guarantees that a public key actually belongs to a certain organization (Fig. -A.122). Specifically, the certification authority provides an electronic certificate which can validate a public key (Fig. -A.123); it also sets time limits during which a certificate may be active. It provides its own encryption and a temporal window in which it can be used. Message authorization code.

Key Management and Encryption without Transmitting Keys

Procedures for secure management of keys remains difficult. The key needs to be delivered to the correct recipient. If the keys are distributed by an insecure channel, they could be stolen. Because of the difficulty of key management, a procedure that creates an encrypted channel without transmitting keys can be useful. The Diffie-Hellman procedure (-A.13.2) can be used to exchange information securely because the keys are never transmitted in the open. This is the principle behind SSH.



Sequence	Description
0.	Certification authority sends private key to HTTP server. This is often done when server site is set up.
1.	User contacts merchant's HTTP server.
2.	HTTP server suggests switching to secure server.
3.	User asks for merchant's public key from certification authority.
4.	Certification authority replies with merchant's public key.
5.	User merchant's contacts secure server.
6.	Secure server responds and user can decrypt page with public key.

Figure A.122: The steps in authentication with a certification authority. (FIG)

Field	Description
Version	Version
Serial number	Unique serial number
Signature	Algorithm used to sign certificate
Issuer	Trusted entity
Validity	Dates for which the certificate is valid
Subject	Name of the certificate holder
SubjectPublicKeyInfo	Algorithms for which the certificate is valid
IssuerUniqueID	ID of trusted entity
SubjectUniqueID	ID of certificate holder
Extensions	Extensions

Figure A.123: The main fields of an electronic certificate (adapted from[3]).

Digital Signatures and Digital Time-Stamps

Hashing is a procedure generally produces an index number from complex number. This unique number can be used as a digital signature. Time-stamps are an application of digital signatures which describe when an information resource was created. An inventor might want to be able to verify the date on which his or her invention was created, or a hospital may want to confirm the time and date when an X-ray of a patient was taken. Secure hashing, which is similar to encryption, generates a unique hash code for the object; this can be widely published so that its time cannot be disputed. Simply including a digitized time in an ordinary encryption is not proof because that time-stamp could have been forged before the encryption. This can be a technique for authentication.

A time-stamp system is based on publishing a rolling hash code (Fig. -A.124). Provides trust (5.2.3). Content encrypted with that key must have been in that sequence based on a reconstruction of the sequence of values. The result is published in a newspaper classified advertisement. Because the newspaper is dated and widely distributed, the time stamps must have been generated on that date.

Encryption Policies

Encryption attempts to scramble messages so thoroughly that they cannot be decoded except by someone with the key. This technology may be abused; it could enable criminals to communicate without any possibility of detection. The U.S. government has attempted to control the distribution of encryp-

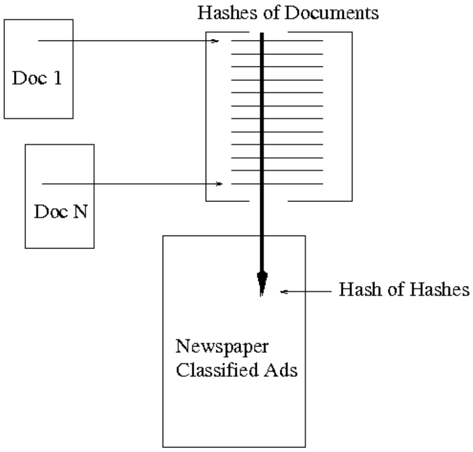


Figure A.124: Steps in time-stamping digital objects. The documents are hashed and the hashes are stored in a list. This list is hashed and the resulting hash code is published in a newspaper. As with any public-key system, the list can be read, but the time-stamp agency can prove that only they could have produced that hash. (FIG)

tion technology by prohibiting its commercial export. Critics of this policy argue that the encryption technology should be freely available. However, that effort has not been generally successful and the debate has shifted to whether there should be a way for government officials to over-ride the encryption in some cases. Privacy advocates disagree with the inclusion of an over-ride capability. Current encryption technology is so good that for all practical purposes it cannot be broken. The most serious problems with credit card authorization on the Web have not been with the algorithms, but with the control of decrypted card numbers that were stored in a database.

A.13.2. Digital Encryption Standard (DES)

DES is the basic procedure for symmetric key encryption. It proceeds through a series of permutation, rotation, using the Boolean XOR operator (3.9.2). This is fairly fast and is reversible, but can be difficult to crack depending on the number of bits in the XOR. A schematic of the steps is shown in Fig. -A.117.

1 10111000	00101110
2 -	-
3 -	-

Figure A.125: Simplified version of DES using eight bits, rotation, and XOR.

A.13.3. Public-Key Encryption Algorithm

Applications of public-key encryption were described earlier (-A.13.1). Here we explain the RSA public-key encryption algorithm following[69]. This is sometimes called a "trapdoor" or "knapsack" algorithm because it is easy to go in one direction but difficult to go in the other direction. This is especially true for very large values. Begin by selecting two prime numbers, p and q and an encryption key, e. Those values can be used to derive the decryption key, d (Eq. -A.20).

$$(d * e)\ mod\ ((p-1)(q-1)) = 1 \tag{-A.20}$$

A message, M, can be encrypted to a cipher, C, with Eq. -A.21. When we want private encryption of messages, e and $p * q$ together can be used as a "private key".

$$C = M^e\ mod\ (p * q) \tag{-A.21}$$

The cipher can be decrypted using Eq. -A.22 based on p, q, and d. While d and $p * q$ may be known, as long as the individual values of p and q are private, it is extremely hard to find e.

$$M = C^d \ mod \ (p * q) \tag{-A.22}$$

As an example, if we took two prime numbers, $p = 11$ and $q = 3$ and we select $e = 13$. d can be calculated from Eq. -A.20.[3]

$$(d * 13) \ mod((11 - 1)(3 - 1)) = (d * 13) \ mod \ (20) \quad = \quad 1 \tag{-A.23}$$

$$d \quad = \quad 17 \tag{-A.24}$$

For instance, this might be the ASCII code a text message. Now, imagine that we want to transmit the number "9" as a message, M. The cipher can be calculated with Eq. -A.21 along with the values of e. [4]

$$C \quad = \quad 9^{13} mod(33) \tag{-A.26}$$

$$C \quad = \quad 15 \tag{-A.27}$$

The receiver can then decode the cipher using the decryption key, d and the value of $p * q$ and Eq. -A.22 to recover the value of the message, M.

$$M \quad = \quad 15^{17} mod(33) \tag{-A.28}$$

$$M \quad = \quad 9 \tag{-A.29}$$

Because the public-key calculations are computationally expensive, an entire message is typically not encrypted with this technique. Rather, the DES algorithm may be used to encrypt the message and only the DES key is encrypted with the public key algorithm.

A.13.4. Public-Key Infrastructure (PKI)

Certification Authorities and Electronic Certificates Fig. -A.123 shows the steps required by a certification authority.

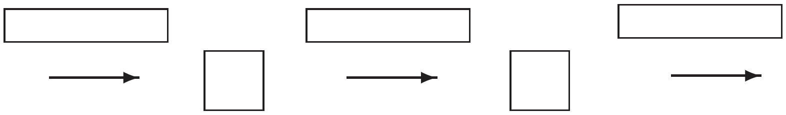


Figure A.126: Stream cipher.

A.13.5. Cryptographic Protocols

Encryption can be the foundation of low-level protocols. For instance, employing middleware to provide anonymity.

The encryption procedures just described can be applied in many ways, Cryptographic that manages different types of interaction. Determining the highest salary among a group of people without being able to identify who has it.

Encryption is possible solution to security rather than a system solution.

"Pseudonym" systems.

[3] For large values, this calculation can be simplified with Euclid's theorem.

[4] Note that even for small values these exponents will overflow most computers. The computation can be made more tractable by decomposing the exponents. For instance:

$$9^{13} mod(33) = [9^6 mod(33) * 9^5 mod(33) * 9^2 mod(33)] \ mod(33) \tag{-A.25}$$

Compared to credit cards, cash provides anonymity because there is no electronic trail.

Secure multiparty communications. Protecting privacy (8.3.1) and data mining. A bank can guarantee a payment without the source of the funds being directly identified. This provides about the same level of anonymity as cash (Fig. -A.127)[23].

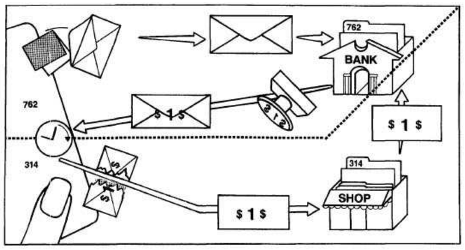


Figure A.127: An example of a cryptographic protocol for ecommerce. In the top panel the user requests a certificate from the bank for a fixed amount. In the bottom panel, the certificate is given to a merchant. (adapted from[23]). (redraw)

A.14. Servers and Networks

We have seen range servers from databases, to Web sites, to repository servers. In middleware (7.7.1). A server is a networked computer that specializes in delivering data and information. Network security. Peer-to-peer.

This does not include the human or organizational issues.

We have touched on servers in many sections.

A.14.1. Database Systems

Database management systems (DBMS) ((sec:dbmsbasic)). Indeed, these may be federated systems in which case they would are distributed databasemanagement systems (DDBMS).

While we emphasized databases for retrieving information, many databases also need to store information received from users.

Unitary transactions. ACID: Atomic, Consistency, Isolation, Durability

CRUD: Create, Read, Update, Delete.

Transaction Management

Nested and distributed transactions.

Database transactions. Lock to make sure the cannot be changed by another process. Prevent conflicts of two disk activities at the same time.

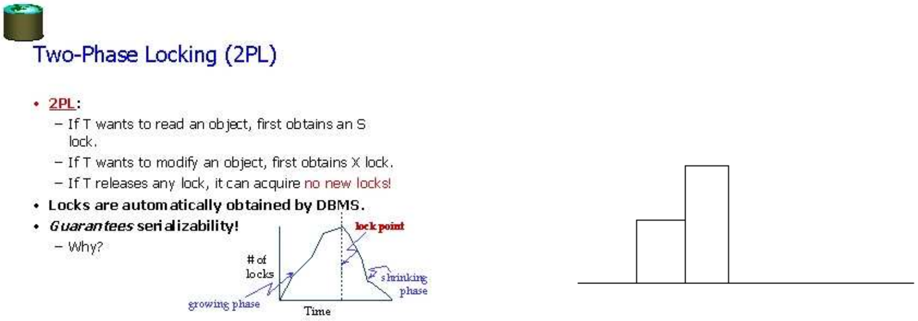
Several transactions may occur simultaneously. Concurrency control. Suppose you are online and browsing for an airplane ticket. You would be very annoyed if you have picked a flight and seat but before you complete the purchase somebody else slips in and purchases that seat. This problem can be helped by creating a lock on the seat one you request it. While the lock in effect nobody else can select that seat.

Avoid conflicts and deadlocks.

Rollback points.

Check-in and check-out to make sure the do not overlap.

Two-phase locking (Fig. -A.128). Growing phase and shrinking phase. Locking phase and release phase.



Two-Phase Locking (2PL)

- 2PL:
  - If T wants to read an object, first obtains an S lock.
  - If T wants to modify an object, first obtains X lock.
  - If T releases any lock, it can acquire no new locks!
- Locks are automatically obtained by DBMS.
- *Guarantees* serializability!
  - Why?

Figure A.128: Two-phase locking. (redraw)

Network Databases

The Web is highly distributed and there are no guarantees as to information being available.

A variety of systems have been developed for managing the coordination of the network infrastructure.

Placement in the Network There is often a tradeoff between disk storage and network usage. Frequently accessed content can be made available. Mirroring.[??]

A.14.2. Web Servers

From Web servers to repository servers and content management systems (CMS) (7.8.0).

Web Server Logs

Servers record a great deal of information about each transaction (Fig. **??**).

Tools for analysis and improved advertising. Able to find IP addresses.

```
208.219.77.29 - - [17/Aug/1999:11:57:58 -0400] “GET /robots.txt HTTP/1.1” 404 207
208.219.77.29 - - [17/Aug/1999:12:01:38 -0400] “GET /snews/ HTTP/1.1” 200 822
208.219.77.29 - - [17/Aug/1999:13:59:46 -0400] “GET /snews/ HTTP/1.1” 200 822
208.219.77.29 - - [17/Aug/1999:14:24:38 -0400] “GET /snews/browse.html HTTP/1.1” 200 665
208.219.77.29 - - [17/Aug/1999:14:36:24 -0400] “GET /snews/form.html HTTP/1.1” 200 1080
208.219.77.29 - - [17/Aug/1999:16:16:51 -0400] “GET /snews/form.html HTTP/1.1” 200 1080
208.219.77.29 - - [17/Aug/1999:16:24:29 -0400] “GET /snews/MDUD/pageImages.html HTTP/1.1” 200 856
208.219.77.29 - - [17/Aug/1999:19:26:07 -0400] “HEAD /snews/MDUD/pageImages.html HTTP/1.1” 200 856
208.219.77.29 - - [17/Aug/1999:19:28:10 -0400] “HEAD /snews/NYBE/pageImages.html HTTP/1.1” 200 425
```

Figure A.129: Web Server log files. Each of part of the Web page to be retrieval such as individual figures is recorded separately.

Anonymizer.com

Caching of Web pages depends on Web usage patterns. It can be on the browser or in the networks; for example at a proxy server.

A.14.3. Link Resolution for Digital Libraries

Some links may be context sensitive. For instance, links for an appropriate copy may depend on contracts. Links in the local context.

One strategy for organizing virtual or distributed collections employs “Digital Object Identifiers” (DOIs). These unique codes are composed of a prefix that describe the directory and publisher, and a suffix that assigns the object a code of the publisher’s choosing.

Manage access rights. Digital object identifiers (DOIs) (-A.14.3).

The “appropriate copy” linking service[55] uses the Handles protocol[40] (-A.14.3). It is an “appropriate copy” in the sense that a contrast or license exists for accessing that content. This is also termed “context-sensitive linking” (Fig. -A.130). From DOIs (-A.14.3).

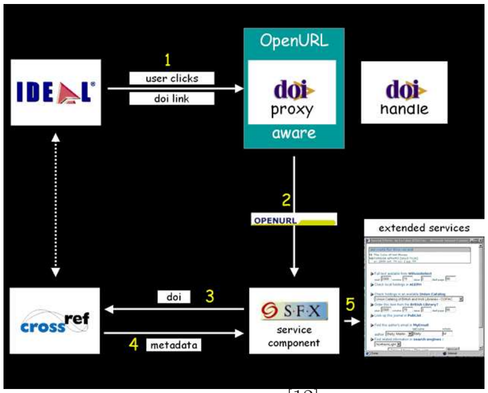


Figure A.130: Context-sensitive linking[12]. (check permission)

A.15. Transmission and Networking

It is not the intention of this text to provide a general course on transmission; rather, we focus on data transmission. Ideally, transmission should be fast, flexible, and real-time. This can greatly affect multimedia presentations.

Figure A.131: Telephone lines in rural Virginia (from LC)

Distributed protocols.

A.15.1. Data Transmission

Data has to get from one place to another. Transmission costs are falling rapidly and increasing in portability.

Digital versus analog. Asymmetric links. The back channel does not necessarily have to be as high bandwidth. Symmetric network, have sources equal to sinks.

Broadcast and Wireless

There are many technologies and many ways of delivering content. In a broadcast transmission, an antenna sends signals into the air; broadcast is widely used by traditional analog radio and video stations. Broadcast normally sends signals to anyone with an appropriate receiver.

Spectrum The electromagnetic spectrum includes radio frequencies used for communication service. Different parts of the spectrum are useful for different applications. These are licensed to avoid conflicts in communication. This licensing regulates, for instance, the number of broadcast television stations in a region. In the U.S. permission in allocated by the Federal Communications Commission (FCC). In some cases, the spectrum is quite valuable and it is generally auctioned to the highest as a public resource. Different parts of the spectrum are suitable for different applications. Fig. -A.132 shows how the spectrum is divided for communications.

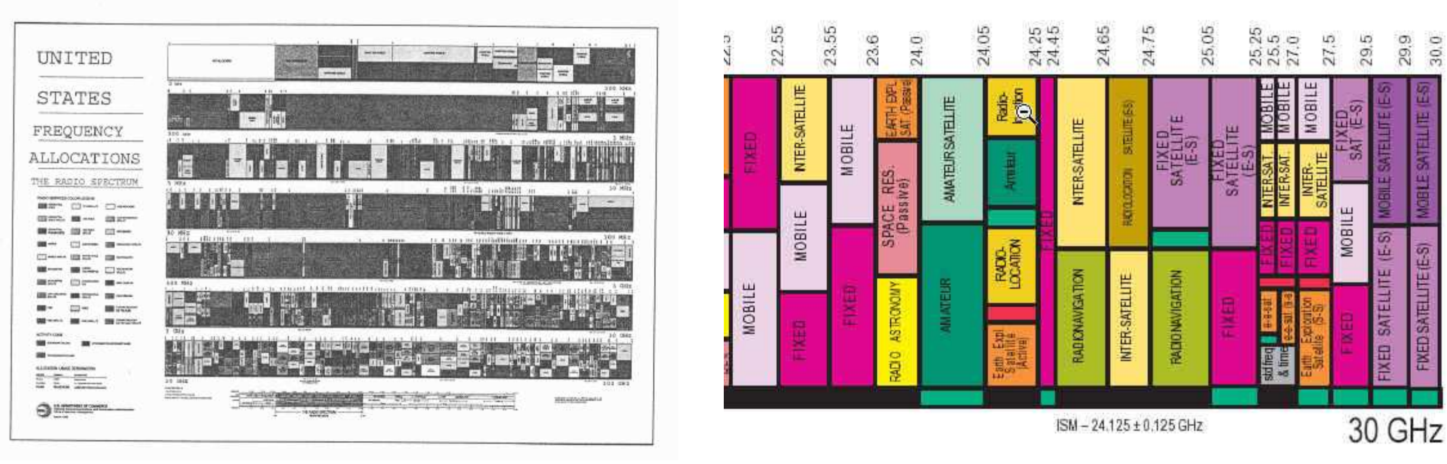


Figure A.132: Chart showing the allocation of electromagnetic spectrum for communication in the U.S. (left) and a detail of the chart (right)[53].

Wireless Wireless transmission allows for portable, and hence nearly ubiquitous, dissemination of information. When broadcast is used for personal communications. Analog versus digital radio versus IR transmission microwave.

Who owns wireless spectrum. Is it a public resource. Commons wireless networks. Sharing bandwidth. For instance, garage door openers share a spectrum with fighter aircraft.

Wireless and mobility of services.

This has the potential to make highly portable services.

Cellular – what is a cell. Microcells. Fig. **??** shows how cells work in a cellular telephony system.

Coordination between cells.

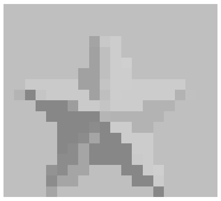


Figure A.133: Cellular telephony.

Multimedia over wireless poses substantial bandwidth difficulties.

Spread Spectrum With traditional radio, the broadcast is on a single frequency. However, it is also possible to spread information across different wavelengths of the spectrum (Fig. -A.134) [**?**].

Satellite Relay Satellites provide communication coverage in remote locations. Several generations of satellites have been deployed. One important difference between them is their orbits. “Remote sensing” satellites. There are two fundamental types of communications satellites, those in Geosynchronous Earth Orbits (GEOs) and in Low Earth Orbit (LEOs). The GEOs stay in one position above the earth. GEOs at X KM (24K miles). footprints. LEOs are not geosynchronous. Several sets of LEOs such as Teledesic and Iridium are being deployed. Delay in satellite communications makes two-way voice links difficult.

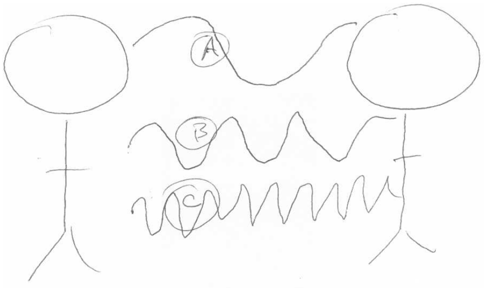


Figure A.134: In spread-spectrum communications, parts of a message are communicated on different wavelengths. Because different wavelengths are used, the message can be robust and difficult to intercept.

Location Technologies

For games and for mobile services. Coupon alert during shopping,

Providing better bus services. How to optimize bus services for times. Auctioning spaces (perhaps by better prices.

Managing location with tradoff of bandwidth and energy use.

Location-related search. Walking routes by mining previous trajectories. Finding a stationary object.

Global Positioning System (GPS) Fig. -A.135 illustrates how a GPS can calculate the position of an object on earth based on the difference in the timing of signals received from the two satellites. Better resolution, including 3-D position, can be obtained by using the signals from three or four satellites.

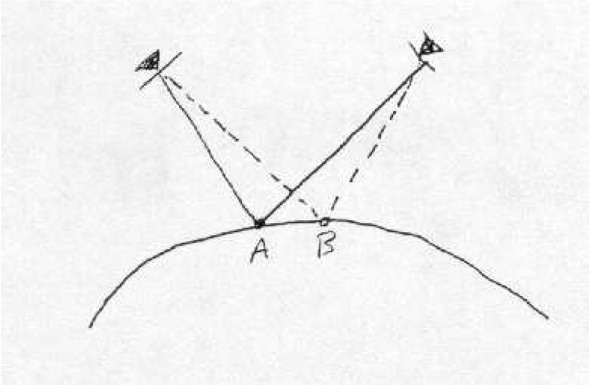


Figure A.135: Global Positioning System (GPS) position is obtained from satellite positions.

Indoor Location Properties of waves. Wave propogation method of location.

Signal strength-maps. Know characteristics of of the signals in a building. Problem of people walking around buildings. Very costly and time-consuming. Privacy problems in all this monitoring.

Navigation Based Position Accuracy Navigation through space. Inertial navigation systems (INS) Can use navigation for Compass, Accelerometer, Gyroscope.

Coordinating locations with camera or image processing.

Understanding the meaningfulness of behavior. Judging a person's intentionality for their motion.

A.15.2. Digital Networking

Digital networking makes distributed information systems possible. The information revolution depends on getting the information there. Internet versus the Web.[??]

Error control in networks.

Packets and Routing

A distributed network is designed can be robust to failure. In a centralized network consider what happens to a failure at the central node.

Packets are really sets of electrical pulses.

Sniffing.

Local Area Networks (LANs) The configuration of the network reflects the Robust networks (Fig. -A.136). This is analogous to social interaction (1.2.1). How best to get from one place to another.

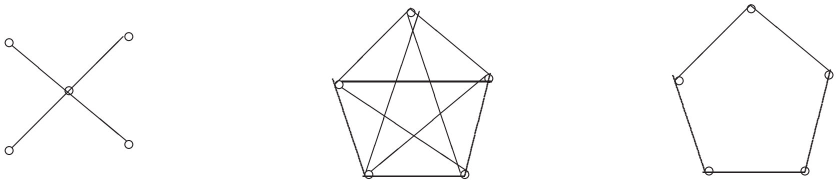


Figure A.136: Three local area networks are illustrated. In a ring (right) the packets will be sure to reach all nodes. The other networks require routing. The distributed network should be more robust to failure than the "star" network.

Ring network.

Circuits and Networks

In traditional telephony, there is an electrical circuit between the mouthpiece of one telephone and the receiver of the other telephone. This discussion focuses on packet networks. A packet is a collection of electrical signals that carries both header information and data. The header information describe the destination of the packet and the type of protocol it follows.

Ethernet is the most common packet network but many others have been proposed. Ethernet is the primary example of CMDA (collision networks) but increasingly it is being adopted by wireless systems. This provides robustness, but not if saturated. Ethernet on a single network. Gigabit Ethernet.[??] VBR (Variable Bit-rate Transmission), ATM layers — physical layer, ATM layer, adaptation layer.

Non-CDMA networks.

Addresses and Ports

In a circuit where several machines are connected, the machines must be given addresses to distinguish them from each other.

Addresses and Domains Typically, networks are interconnected. To go beyond a local network requires gateways for routing to other networks. IP and Class B, C, D address,

There are many policy issues surrounding internet naming [**?**].

Packet Protocols

A protocol is a standard for communication to ensure that a transmission goes to the right place. There are the IP-level protocols.[??] Service protocols, such as http, are discussed above ((sec:http)).

The Internet is a harsh environment for packets (-A.15.2). If key data is in only a few packets and those are lost due to congestion, serious problems can ensue. Often an adaptation is made to network environments by dropping frames. In current implementation, all packets are given equal priority.

A.15.3. Multimedia and Hypermedia and Networking

Multimedia networking has special requirements. Even small delays can make a difference in transmissions, so scalability is important.

Networking and Special Effects

Where in network are special effects completed? Reconstruct fades later locally [?]. Real-time interaction.

Network-Scalable Multimedia Services

A traditional video stream has fixed-rate bit streams. However, interactive multimedia services are often "bursty" (Fig. ??).

A.15.4. Audio Delivery

Because it has lower bandwidth requirements, audio services are easier to develop in the short term than are video services.

Internet telephony. This can mean many things to many people. ITU H.323. The problem of congestion along routes can be a significant factor. However, many internet telephony services run P2P protocols.

The telephone is a real-time multimedia service. Indeed, the real-time restrictions are stricter here than for delivery of audio or video; very little delay can be tolerated.

Repair of audio, In many cases, audio is fairly predictable. Thus, if a packet and its data are lost a good guess can be made about how to replace it.

Communication services such as live telephony have very stringent network requirements. VOIP, voice over IP.

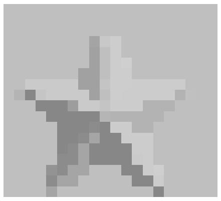


Figure A.137: VOIP.

A.15.5. Video Delivery

S-Video, Composite video[??]

Video Broadcast and Networking

Video is not one technology but many. There is a fundamental distinction between analog and digital transmission. You are probably most familiar with analog video, which is broadcast or delivered by cable to your television. Most new video technologies are digital. Digital video allows pictures to be computer-processed. Special effects can be generated and frame rates and compression can be easily controlled. Digital video generally requires very large amounts of data compared to images and even audio. Low-level networking issues and video hardware are discussed in -A.18.0.

Digital video also allows for delivery of video by packet networks. In the near future, broadcast quality video is not likely to be carried on the Internet because of the large amounts of data involved. As gigabit networks and satellite delivery are more widely deployed, this may become common.

Analog Broadcast Video

Broadcast television started as black-and-white.

As color television was developed, it was necessary to allow the large number of existing black-and-white television sets to be able to received programs transmitted in color and to allow color television sets to receive black-and-white programs. Thus, the color signal was superimposed on the black-and-white signal. SMPTE (Society of Motion Picture and Television Engineers) standards were established to do this, and to handle the special cases of color superimposed

on an analog broadcast signal.

There are two widely used broadcast formats: NTSC and CECAM-PAL. NTSC is used in North America while CECAM-PAL is used in most of the rest of the world.

High-definition television (HDTV) is a widely discussed standard.

Digital Video

There are many ways of transmitting data by wire. In addition, digital video can be processed in other ways. Video on demand is one service that can be provided with this technology.

Digital video is delivered over the network or by wireless. Digital Video Broadcast (DVB) can be of higher quality than analog. The ATSC (Advanced Television Standards Committee) establishes criteria for DVB.

In streaming video, frames are sent and viewed consecutively as they arrive. Streaming may be more efficient; viewing of a video can start sooner because one does not have to wait for an entire file to download. One limitation with streaming is that there may be congestion in the network and some frames may arrive late or not at all. A second limitation is that streaming video is usually unicast, that is, only one client is connected to the server at one time. Multicasting allows many people to be served by a single a video source while minimizing network load.

Combining video with many other services.

Multicasting may also be used for other services such as distribution of audio and games.

A.16. The Internet

The Internet is the international collection of packet networks which implements the Internet Protocol (IP). It was designed as a distributed network to promote robustness and survivability. During the 1980's private networks grew but many of these used proprietary protocols and were interconnected.

The Physical Internet While we have focused on protocols, but, of course, the Internet is made up of communication lines and routers. Avoiding congestion (-A.16.0). Map of the Internet

While it is relatively easy to provide high speed network connectivity on major trunks. Feeding that connectivity out to individual locations. Last-mile problem.

Layers of Service Layering is a good strategy for managing complexity (7.7.1). Fig. -A.138 shows the ISO Open System Interconnect (OSI) layers for services. Layering for separating the complexity (7.7.1). This specification is focused on the network and not on the services. Ideally, the layers should be independent of each other. Reference model for how a network should be built.

Layer	**Description**	**Example**
1	Application	
2	Presentation	
3	Session	Circuit connection
4	Transport	TCP
5	Network	IP
6	Data Link	binary data
7	Physical	cables

Figure A.138: The ISO OSI 7-layer model. Each layer is designed to operate separately from the others.

Screendump of router se up.

A.16.1. Internet Economics and Policies

Information networks are embedded into the social fabric.

Economics (8.7.0). Network economics.

Moving computing or moving data.

Packets and network economics[76]. Impracticality of charging per packet. Peering.

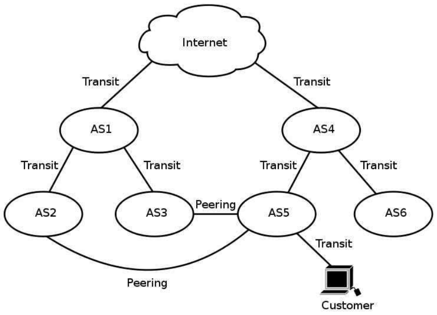


Figure A.139: Internet peering. (redraw) (check permission)

Boundary gateway protocol (BGP).

Internet structure is described in (-A.16.0). The flat-rate business model was instrumental in generating business. Business models for the Internet. Measuring traffic. Pricing. Net neutrality. Dark networks.

A.16.2. Regulating the Internet

Controlling Internet activities by national laws[33].

Government regulation. Cross-border regulation.

Real-Time Services on the Internet

The Internet transmits packets but they may be delayed or destroyed. For instance, if too many packets arrive at a switch at a given time, the buffer may overflow and some of the packets might be discarded. Even if they are not discarded, they may be delayed.

For email, these delays are not significant but for real-time interaction, only minimal delays can be tolerated. While the Web is mostly text and images, as we have seen throughout this book, multimedia is constantly increasing.

A variety of new Internet services have been proposed, including IPv6. Charging and wireless.

Buffering of transmitted information[??]

Real-Time Protocols A variety of protocols for real-time IP services have been proposed. UDP packets are sometimes preferred for multimedia because of the speed.

Robust IP Multicast.

Quality of Service Guarantees The Internet is highly distributed and has many bottlenecks.

Quality of Service (QoS) guarantees and multimedia (blocking and latency). Requires cooperation from routers.

Problem of congestion from packets on the network.

Alternatives to real-time Internet delivery,[??]

A.17. Computing Architectures and Operations

A.17.1. Theory of Computation

What is the way to organize components such as switches and memory to do complex computation. Turing[74].

Interactive computing[5].

The basic units of a Von Neuman computer [77]. 1. Arithmetic unit, 2. Memory 3. Control 4. Input/Output (Fig. -A.140).

Instruction space:	Memory space:
x=x+1;	x
z=x+y;	y
	z

Figure A.140: Stored programs need both instruction-memory space and a data-memory space.

A.17.2. Computer Programming Languages

Effectiveness.

Machine Language

Instruction space and data space.

Formal Properties of Programming Languages

One attribute of a programming language is the ability to express complex material. The ability to do any type of computation is known as being "Turing complete".

Useful for applying algorithms for completing certain tasks.

Formal languages (6.5.2). Parsing and compiling,

A.17.3. CPU Architectures

The complexity of the algorithms wired directly into a CPU chip affects its size, speed, and the heat it generates. Thus, chip designers have two approaches.

Booleans are the basis of the gates used in digital logic (Fig. -A.141). (3.9.2, -A.7.1).

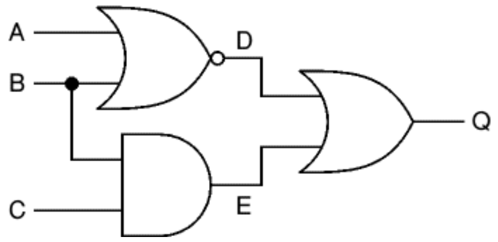


Figure A.141: Logic circuits. The OR gates and the AND gates. (redraw)

A computer may be called on to do a wide range of calculations. When designing CPUs, there was a tendency introduce instructions for and many of those computations as possible. This resulted in so-called Complex Instruction Set (CISC) chips.

However, the CISC chips were more difficult to manufacture, were more specialized, and consumed more heat when operating. Thus, the chip makers decided it was better to simplify the number of instructions. This resulted in Reduced Instruction Set (RISC) chips.

Reconfigurable computing.

Graphics computing. Cell processor.

A.17.4. Distributed Problem Solving

A.17.5. Parallel Computing

A distributed system has several computer processors connected by a network while the network connections are fairly fast, they are not nearly as fast interconnected systems with a shared bus. These centrally connected computers are call "parallel". There are many ways they can be inter-connected. Fig. -A.142 shows multiple streams with crossovers. In this configuration, the results for each stage are passed to all processors active in the second stage. "pipeline" model.

Mesh networks.

Cell computing.

Coordination and computation.

Parallel algorithms.

For some other problems, arbitrary exchanges between processors (Fig. -A.143). There are several different architectures.

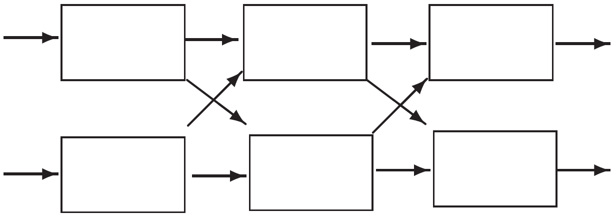


Figure A.142: A parallel computer has multiple connected CPUs.

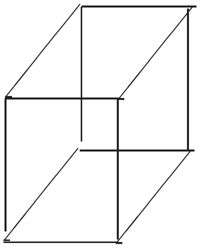


Figure A.143: A cube architecture allows the shortest path for communication among the nodes. We can easily visualize a 3-D cube but it is also easy to wire nodes into higher-dimensional cube, "hypercube," architectures.

Specific algorithms can match these architectures.

Multicore processors.

A.17.6. Grid Computing

We briefly considered grid computing (7.8.1).

Networking, storage, and concurrency.

Large Scale Distributed Storage

Delay across storage.

BigTable.

Storage resource broker[13]. Rule specification. iRODS.

Peer to peer system for storage. LOCKSS Preventing groups from. Grid computing, the storage resource broker[14].

“The SDSC Storage Resource Broker (SRB) is client-server middleware that provides a uniform interface for connecting to heterogeneous data resources over a network and accessing replicated data sets. SRB, in conjunction with the Metadata Catalog (MCAT), provides a way to access data sets and resources based on their attributes and/or logical names rather than their names or physical locations”. QUOTE

Move data around the net based on cluster analysis of how it is used. Importance of keeping a single master copy. Data storage (Fig. **??**). Difficulty of updates.

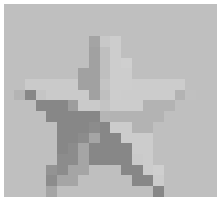


Figure A.144: BigTable.

A.17.7. Models of Computation

Blackboards.

Neural networks (-A.11.4).

Autonomic Computing

Get the system to optimize itself. Self-aware, self-healing[2].

A.18. Input/Output Devices

Although digital processing is increasingly important, it is often necessary to understand the effects of physical processes. Input/output devices.

A.18.1. Audio Devices

Transducer for audio[??]

Microphones

A microphone converts sound in air to electric signals. Directional microphones. Cone of sensitivity.

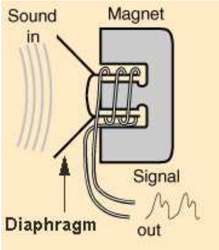


Figure A.145: Microphone. (re-draw-K)

Speakers

Speakers create pressure waves from electrical signals; a speaker's sound box can provide resonance. Different speakers are used for different pitches.

Specialized Audio Processing A-to-D, Digital Signal Processing (DSP) chips![??]

A.18.2. Visual and Video Devices

How to capture and present an array of signals. Here, we briefly survey several technologies. Digital cinema.

Printing Technology

Paper and ink. E-Ink figure. (Fig. -A

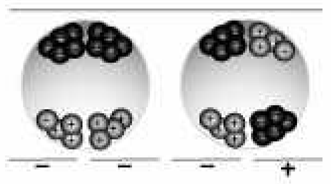


Figure A.146: Black-and-white balls with different electrostatic charges are placed in a clear larger ball. Applying an external charge causes the balls to separate[4]. (check permission)

Cameras and Scanning

Charge coupled devices (CCDs) — solid state cameras. Scanning is the usual approach for digitizing a paper document or picture. It may be done at different resolutions; after scanning, the bitmap can be compressed. Once scanned, images can be archived, distributed, or processed.

Scanners or digital cameras digitize and analyze small areas of a picture and measure the brightness or colors in that small area.

Effectiveness for reproducing readable text. The quality of the scanning for the resolution is shown with the "quality index" (Eq. -A.30). [42]

$$Quality\ Index = h * dpi; \qquad (\text{-A.30})$$

For fragile materials, it is necessary to employ non-destructive scanning.

Video Displays

Refresh rate. Number of pixels on a standard television display![??] The "aspect ratio" of a video display is that of length to width. Larger over smaller. Fig. -A.147 contrasts the aspect ratio of television (A, B) with the aspect ratio for cinema (C).

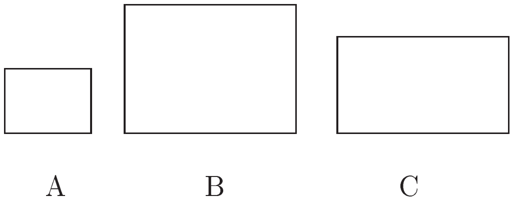


Figure A.147: The aspect ratio is the ratio of the width of a display to its height. The ratio remains constant although the absolute size may change. The ratio 4-to-3, as shown in Panels A and B is the standard for video. While the ratio 16-to-9, as shown in C is used for cinema.

LCD displays, plasma displays, interlacing. The picture is presented on the screen with rasters. Vector graphics. Raster.

Broadcast video. Vertical blanking interval. Difference in frame rates, color depth, etc.

Readability of displays (10.3.1)[34].

Rather than creating a sharp boundary at the edge of a character, which often appears as jaggies, anti-aliasing makes the boundary with a gradual fade to gray. Fig. -A.148 shows anti-aliasing.

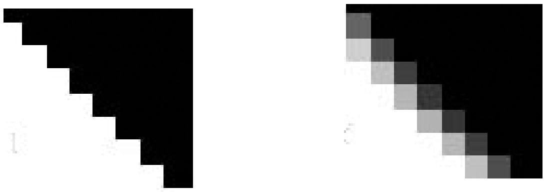


Figure A.148: Rather, than a sharp edge, pixels in a display create a jagged edge (left). To create the appearance of a smoother edge, the edge pixels are grayed out.

Pen tiling of display colors.

BitMap Displays

(Fig. -A.149)

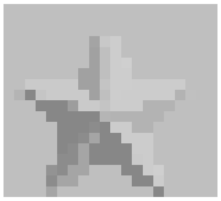


Figure A.149: Memory management for bit-map displays.

Overlapping Windows.

Other Types of Displays

Stereoscopic 3-D A variety of technologies have been developed to make stereoscopic 3-D presentations.

Head-Mounted Displays and Head Tracking Head tracking.

Technologies for Personalized Displays Retinal painting[??]

Immersive Display Technologies

Displays cannot have the same degree of fidelity as reality. In one study of a virtual reality system[56], the display was 0x120, 93-degrees by 61-degrees.

Volumetric Display

Volumetric displays[16]. Painting into plasma.

Printing Technology

Resolution, DPI (dots per inch)[??]

CMYK color, a variation of RGB (-A.2.3), is used for printing.

3-D Hard Copy [1]

A.19. Sensor Technology

Sensors are, typically, simple devices which detect attributes of a system's environment. Attributes such as motion, sound, temperature, air quality, and light are all easy to monitor. Bio sensors. Like you own eyes or ears, typically, sensors have relatively little complex processing capability of their own.

A.19.1. Sensor Devices

Sensor detect properties of the physical world. There are many types of sensors such as body sensors.

Transducers.

Bar Codes

Low cost way to detect portable objects. Laser scanning and reflection. Bar codes (Fig. -A.150). The spacing of the lines. A space can represent a binary code. There are different coding systems. One common system is The Universal Product Code (UPC) was developed to identify production. QR codes. Near-field communication (NFC).

Figure A.150: Bar codes represent numbers with a binary code.

Radio Frequency Identification (RFID)

Passive chip sensor which responds to an external field with coded information (Fig. -A.151).

Near-field communication. Affected by interference.

EPC - electronic product codes.

RFIDs have many applications. Use of RFIDs in hospitals in order to locate patients. Threats to privacy.

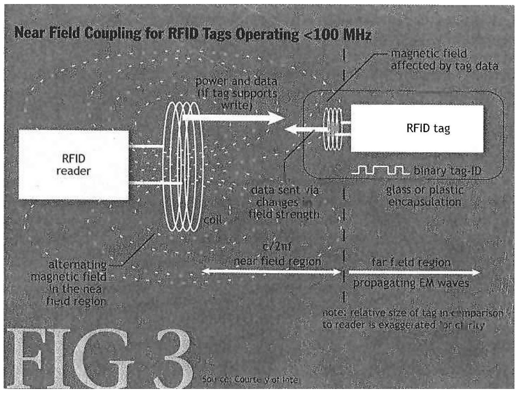

Near Field Coupling for RFID Tags Operating <100 MHz
magnetic field affected by tag data
power and data (if tag supports write)
RFID tag
RFID reader
binary tag-ID
glass or plastic encapsulation
data sent via changes in field strength
coil
c/2πf
near field region
far field region
propagating EM waves
alternating magnetic field in the nea field region
note: relative size of tag in comparison to reader is exaggerated for clarity
FIG 3
Source: Courtesy of Inte

Figure A.151: How an RFID sensor works[10]. (check permission)

A.19.2. Sensor Networks and Sensor Fusion

Sensors can be connected in networks.

Typically they have simple processors, limited memory capacity, and limited power.

Here we consider two specific systems. The signals from these sensors are processed by sensor fusion (-A.19.2). Hierarchical fusion versus mesh or grid fusion. Fusion of similar data versus fusion of dissimilar data.

This often means that low-level information is processed within the network. Sensor fusion combines information from many sources (Fig. -A.152). There is a challenge about how to weight the information appropriately.

Hierarchical sensor networks and communication in sensor networks.

Many applications: Sensor fusion for emergency room data. RFID (-A.19.1).

Generating too much data. We need to automatically filter the data. We attend to (4.2.2) to significant information.

This is often noisy information with ambiguity. During the Cold War, the U.S. Navy maintained an array of sensors in the North Pacific. These sensors had to be able to distinguish submarines from whales swimming in the ocean.

Distributed decision making.

Scientific instruments and data storage.

Privacy issues from potentially invasive sensor networks.

One approach is hierarchical summarization Fig. -A.152.

There can be local interactions among the units such as excitation or inhibition of neighbors. Sensors and feedback. Parallel computing (-A.17.5).

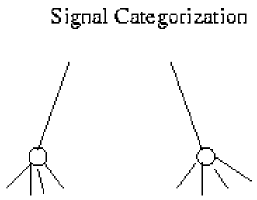


Figure A.152: Data from many sources needs to be combined. Hierarchical organization of sensors.

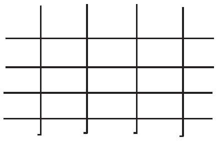


Figure A.153: A "sensor grid" is composed of sensors arranged on a grid. The first level of processing can be communicating and combining evidence with neighbors.

A.20. Storage Technology

A.20.1. Storage Media

Magnetic Storage

Helical scan video. Many formats[??]

Iron-oxide. heads.

There are many media for storing digital information; here, we consider magnetic tapes, magnetic

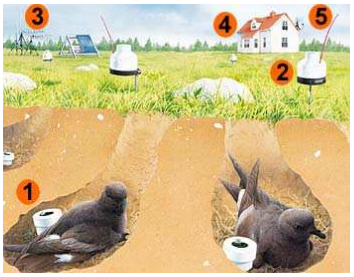


Figure A.154: Ecological sensor network in Duck Island Maine. (check permission)

disks and optical storage. Storage systems must be reliable. Although many technologies have been developed, magnetic disks are so widely deployed that they are hard to beat.

The oxide on recording tapes has a lifetime of about 10 years, after which it becomes unstable. Old tapes may be baked before being played which causes the oxide to adhere.

Many television programs are mastered on film to better preserve them.

Optical Storage

Lasers allow fine resolutions of data to be made on a metallic surface, as, for example, on a CD.

There is about 650 MB of space on a CDROM. This is about 68 minutes of sound recordings at constant bit rate. 16-bit encoding on a spiral track![??]

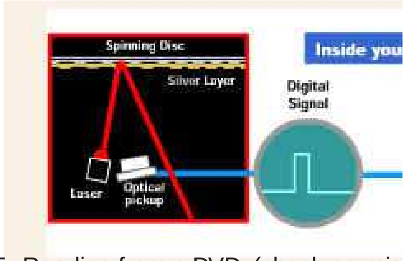


Figure A.155: Reading from a DVD. (check permission)

The digital video disk (DVD) can store 4.4 GB per disk. It has a double layer of reflective material and is double sided. It also improves the density of storage by using a blue laser for reading the disk rather than the red lasers used by the CDROM.

A.20.2. Low-Level Data Storage

Parity and Check Sum

Check that the data has not be corrupted when it is transmitted on a network or stored on a disk. Checksum for credit card verification. Fig. -A.156. If any of the bits in the data have been corrupted a recalculation of the parity bit may flag the problem.

Error-correcting codes.

Data	Parity bit
0 1 0 0 1 1 0 1	0
1 1 1 0 0 1 1 0	1

Figure A.156: A parity bit is calculated as a count of the number of even or odd bits.

Placement of Content on Disk Drives

There are physical restrictions on how data can be placed on a disk. The data must be placed on tracks, and the heads must be positioned above those tracks in order to read the content.

When multimedia content are stored on a disk.

Striping, Speed of streaming.

Disk caching. Random positions are better than standard placement.

Archival Storage

All physical storage media are unreliable. We want to be sure that one reliable copy of a document is preserved. Digital preservation earlier (7.5.1). How to be sure that the originals are not able to be easily corrupted. LOCKSS protocol (Fig. -A.157). When numerous sites are polled, they can essentially take a vote to determine whether any of the copies has been corrupted. If a corrupted file is found, the good version can replace it.

Figure A.157: In the LOCKSS protocol, a target version of a document can request that a comparison be made with other stored versions of the same document[59]. If a discrepancy is found a voting procedure determine which copy has, most likely been corrupted. (redraw)(check permission)

Related Books

- Cheswick, W.R, Bellovin, S.M., and Rubin, A.V. *Firewalls and Internet Security: Repelling the Wily Hacker.* Addison Wesley, Reading MA, 2003.
- Copi, I.M. *Introduction to Logic.* Macmillan, New York, 1982.
- Duda, R.O., Hart, P.E., and Stork, D.G. *Pattern Classification.* Wiley, New York, 2000.
- Flake, G. *Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems, and Adaptation.* MIT Press, Cambridge MA, 1999.
- Gazzaniga, M. *Human: The Science behind What Makes us Unique.* Ecco (Harper-Collins), New York, 2008.
- Gell-Mann, M. *The Quark and the Jaguar, Adventures in the Simple and the Complex.* W. H. Freeman and Company, New York, 1994.
- Goldsmith, J., and Hu, T. *Who Controls the Internet: Illusions of a Borderless World.* Oxford University Press, New York, 2006.
- Singh, S. *The Code Book.* Anchor Books, New York, 1999.
- Wilson, R.J. *Introduction to Graph Theory.* 4^{nd} ed. Prentice Hall, New York, 1996.
- Lewis, T.G. *Network Science: Theory and Application*
- Witten, I.H., Moffat, A., and Bell, T.C. *Managing Gigabytes: Compressing and Indexing Documents and Images.* 2^{nd} ed. Morgan Kaufmann, San Francisco, 1999.