Information Theory, Semantics

INSC 702: Advanced Topics in Information Science

Shimelis Assefa

Starting Points

- Information Theory, MTC
- Information Measure
- Informational Content
- Semantic Theory of Information, KFI
- Shannon Vs. Dretske (MTC Vs. KFI)

Outline

- Information Theory
  - Amounts of Information, Measures of Information
  - Information Sources
  - Entropy
- Semantic Theory of Information
  - Information & Knowledge
  - Truth & Meaning
  - Principles of KFI
- Contribution to Information Science

Information Theory

- The mathematical theory of information.
- Claude Shannon (July 1948) – A mathematical theory of communication.
- Origins of the theory – from research on – coding and transmission of information.
- Other engineering problems –
  - time and cost effectiveness in Signal transmission,
  - search for suitable codes compatible with the communication channel,
  - the design of noiseless channels.

Information…

- In Shannon's own words, the problem which motivated his research is described as follows:
  - The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point.
  - Frequently the messages have *meaning; that is, they refer to or are correlated according to some system with certain physical or conceptual entities.*

Information…

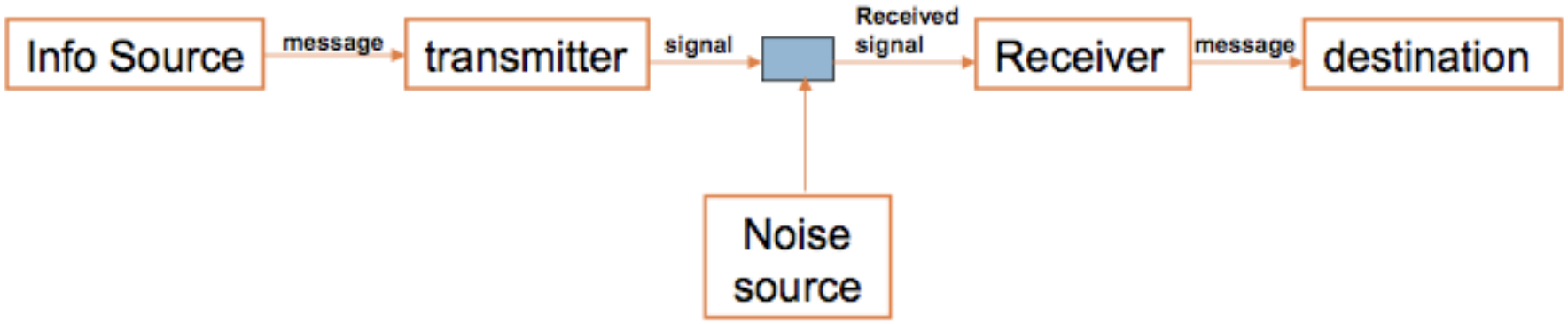
- In Shannon's…
  - *These semantic aspects of communication are irrelevant to the engineering problem.*
  - *The significant aspect is that the actual message is one selected from a set of possible messages.*
  - *The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.*

Information…

- Shannon's concern was the fidelity of telecommunications signals.
- Radio waves, and the determination of the effective capacities of telecommunication channels.
- The measurement of information, the amount of information, not on the informational content (semantics).
- The semantic aspects of communication are irrelevant to the engineering aspects.
- This does not mean that the engineering aspects are necessarily irrelevant for semantic aspects.

Information…

- A model of the communication process (after Shannon & Weaver, 1949).
- In signal transmission, noise is the vehicle for the effects of entropy, i.e., noise degrades the signal to some degree.



Information…

- Information theory or MTC has two primary goals:
  - The development of the fundamental theoretical limits on the achievable performance when communicating a given information source over a given communications channel using coding schemes from within a prescribed class.
  - The development of coding schemes that provide performance that is reasonably good in comparison with the optimal performance given by the theory.

Information…

- It was in this sense that Shannon developed his Information Theory as a model of communication.
- Communication processes-which involve objects and events as physical information carriers (specifically sign vehicles).
- In the context of sign-based information processes, Information Theory can be more generally viewed as a theory focused on the syntactic dimension of these processes (coding).
- Since these sign vehicles are empirical prerequisites for information processes to occur. it is quite appropriate to interpret Shannon's Information Theory in the context of sign processes as a theory of ***information potential*** of sign vehicles.

Amount of Information

- The mathematical theory of communication (MTC) is quantitative.
- Measures the amount of information that may be generated by an event or a realization of a specific state of affairs.
- The theory tells us how much information a signal carries, but not what information is carried.
- MTC offers two communicational constraints essential to any theory that aims to provide an understanding of informational content:

Amount…

1. MTC indicates both the type of signal appropriate for transmitting a specific informational content and the type of informational content that can be transmitted.
2. The mathematical constraints are of considerable help in designing a semantic theory of information, but to do so at this stage will require a different set of theoretical knapsack, may be KFI is the one.

Amount…

- Shannon did establish this connection by making certain assumptions about the relationship of probability as a measurable physical quantity on the one hand and the phenomenon of information on the other.
- Essentially, this connection is based on the assumption that information is not directly observable and thus not directly measurable, but can be observed only in terms of associated phenomena, some of which may be measurable.

Amount…

- One such associated observable, Shannon claimed, is the probability of an event:
  - the more improbable an event is, the more information is contained in a message that the event did happen.
- Thus, although probability may not be the same thing as information, it might, perhaps, be a sufficiently good ***indicant*** of it**.**
- For example, the state of a quicksilver column in a thermometer, observed in terms of its extension or height is not the same thing as temperature**.**

Amount…

- But the state of a quicksilver column in an appropriately constructed device is a good indicator or index of temperature.
- We can determine the temperature of an object by bringing it into contact with a thermometer and measuring the height of the quicksilver column in the thermometer.
- The analogy here is - why not measure information by measuring the associated probabilities in some appropriate fashion?

Amount…

- Probability as an index of information (roughly speaking) can be equated with the contents of a message.
- Probability is more specifically interpreted and related to the function h(p,), called information measure of an event, in the following senses:
- (1) The smaller is the probability of an event to occur, the greater is the “amount of information” conveyed by a message that the event did in fact occur, i.e. Pi > Pi implies ***h(Pi) < h(Pj).***

Amount…

- (2) There is no information contained in a message conveying the occurrence of a certain event, i.e. pi = 1 implies h(Pi) = 0.
- (3) If two events are independent of each other, then the amount of information conveyed by a message that both these events occurred is equal to the sum of information contained in messages conveying the occurrence of these two events separately,
  - i.e. if the event ***E is a joint event of two independent events Ei and Ej, then***
  - ***P(E) X P(Ei)P(Ej) implies h(p) = h(pi) + h(pj).***

Amount…

- In general, info theory identifies the amount of info associated with, or generated by, the occurrence of an event (or the realization of a SOA) with
  - The reduction in uncertainty
  - The elimination of possibilities
- The general formula for computing the amount of info generated by the reduction of n possibilities (all equally likely) to 1 is: $I(s) = \log n$
- ***S = Source; I(s) – the amount of info associated with, or generated by S.***

Information Sources

- Information contexts are shaped by info sources.
- Info sources are defined as sets of events (or states of affairs) with certain probability of occurring.
- Each event that forms a part of that source is considered as a “source signal”, or “source symbol.”
- MTC measures the quantity of information that is transmitted from one point to another.
- “How much information there is at point r about what is transpiring at s”.
- It is concerned with the statistical properties of the ‘channel’ connecting r and s.

Sources…

- Other important concepts in information theory are entropy, noise, redundancy, and feedback.
- Information theory is closely related to cybernetics and system theory.
- Information theory gives rise to certain understandings of information such as the view that information is something modular that may be
  - divided in discrete units and
  - that information is something flowing in a system and that this "something" can be measured, processed and in varying degrees automated and controlled.

Entropy

- Shannon's definition of information is based on the notion of entropy.
- MTC is built around the paradigm of entropy.
- A measure of the degree of disorganization in a system which reflected a tendency for any state of affairs to lose order and become random.
- The entropy is defined by:
  - H(P1, P2, …, Pn) ≡ $\sum_{i=1}^{n} P_i \log_a \frac{1}{P_i}$

Entropy…

- In short, the quantity of information generated at the source ***S*** is the entropy of the source**.**
- The average amount of information generated at the source ***S***.

Semantic Theory of Information

- Founded on Shannon and Weaver's MTC
- Developed and elaborated into:
  - a cognitive, functionalistic theory,
  - individually oriented, and deals with the content of information.
- The topics are: the information process from perception to cognition, and how concept formation takes place in terms of digitization.
- Other important issues are the concepts of information and knowledge, truth and meaning.

Semantic…

- Dretske bridges a gap in his attempts to connect MTC with semantics.
- Dretske gives an answer to the question of how physical signals become meaning in the mind of a receiving subject.
- For Dretske semantic theory of information denotes that the amount of information in a statement is based on **logical alternative** elements of contents that are ruled out in the statement.
- It is close to Shannon and Weaver's concept of information, but is built on **logical principles** and not on **probability theory.**

Semantic…

- Dretske considers truth to be a key concept: there is a connection between information and truth.
- Information in its nature is ‘true’.
- Dretske wants to apply the objectivity of the mathematical information theory to questions of information content.
- It is an objective theory, independent of language and ‘species’.

Semantic…

- For Dretske information is related to communication.
- As representation, information and communication belong together.
- Communication has an objective; it is controlled by intentions.
- To understand an utterance, you must know the speaker’s intention.

Semantic…

- It is intentionality which connects thought and object –the mind and the external reality.
- Information is natural, objective commodity, independent in its generation and transmission, of interpretative processes.
- Objective ‘commodity’ not far from Buckland’s ‘information as- thing’.
- Information is objective “…the sort of thing that can be delivered to, processed by, and transmitted from instruments, gauges, computers, and neurons…it exists whether or not anyone appreciates it or knows how to extract it.”

Semantic…

- Dretske develops information theory to explain various fundamental intentional notions -knowledge, belief, perception, conceptualization and meaning.

Information and Knowledge

- To Dretske information is the principal idea of (nearly) any mental activity.
- Dretske is attempting to develop a theory of information that will explain the notion of semantic content, or the propositional content of a signal.
- Speaking of information, Dretske says ‘information is an objective commodity, something whose generation, transmission, & reception do not require or in any way presuppose interpretive processes. ‘

Information and…

- He considers what is out there, hovering over the water, as plain and pure information.
- He starts *Knowledge and the flow of information (KFI) with a biblical allusion: 'In the beginning there was information. The word came later. The transition was achieved by the development of organisms with the capacity for selectively exploiting this information in order to survive and perpetuate their kind.'*

Information and…

- Knowledge–information produced belief.
- Perception–a process in which incoming information is coded in analog form in preparation for further selective processing by cognitive (conceptual) centers.
- Belief–the thinking that something is so. It is characterized in terms of the instantiation of neural structures that have, through learning, acquired a certain information-carrying role.
- Digitalization: “a process whereby a piece of information is taken from a richer matrix of information in the sensory representation (where it is held in ‘analog’ form) and featured to the exclusion of all else.”

Truth and Meaning

- An important characteristic of Dretske's concept of information is that information must be true in order to produce knowledge.
- But truth and meaning should not be mixed up.
- Knowledge is based on true information.
- From Dretske's point of view meaning (but not truth) is a relative concept which can be a little hard to understand.
- Meaning, and the constellation of mental attitudes that exhibit it, are manufactured products.

Truth and…

- The raw material is information.
- Meaning, as a feature of psychological states, does presuppose cognitive activity.
- It evolves from the flow of information through, and the adaptive requirements of, simple physical information processors.

Principles of KFI

- Some of Dretske's principles
  - If A carries the information that B, and B carries the information that C, then A carries the information that C (Xerox Principle)
  - If a signal carries the information that s is F, then,
    - (a) the signal carries as much information about s as would be generated by s'sbeing F,
    - (b) s is F, and
    - (c) the quantity of information the signal carries about s is or includes that quantity generated by s'sbeing F (and not, say, by s'sbeing G) (characteristic of information)

Principles…

- Principles…
  - A signal, ***r***, carries the information that ***s*** is F = The conditional probability of s's being F, given ***r*** (and what the recipient (k) of the signal knows), is 1 (but, given only what the recipient knows, or K alone, less than 1) (informational content of a signal)
  - K knows that ***s*** is F = K's belief that ***s*** is F is caused (or causally sustained by) the information that ***s*** is F (Dretske's information based theory of knowledge)

Principles…

- Principles …
- A signal carries the information that s is F in digital form iff
  - the signal carries no additional information about s, no information that is not already nested in s's being F.
  - if the signal does carry additional information about s, information that is not nested in s's being F, then the signal carries this information in analogue form (addressing the role of sensation and perception in cognition, Dretske distinguishes between analogue and digital forms of encoding information)

Contribution to IS

- IR - the information science field concerns itself with modeling the world of publications, with a practical goal of being able to deliver their contents to inquirers on demand'.
- This statement represents the tendency to focus on the micro level of analyses, for example in information-searching behavior.
- Dretske's explanations regarding concepts such as concept formation, information and knowledge provide a contribution to the understanding of information science at a macro level (the metatheories), hopefully with implications for the micro level (practical implications).

Contribution…

- The roots of the discipline of information science lay in post-WWII developments, such as:
  - the Shannon-Weaver information theory model,
  - Norbert Wiener's conception of the science of cybernetics, and
  - rapid advances in the design and production of electronic computers.

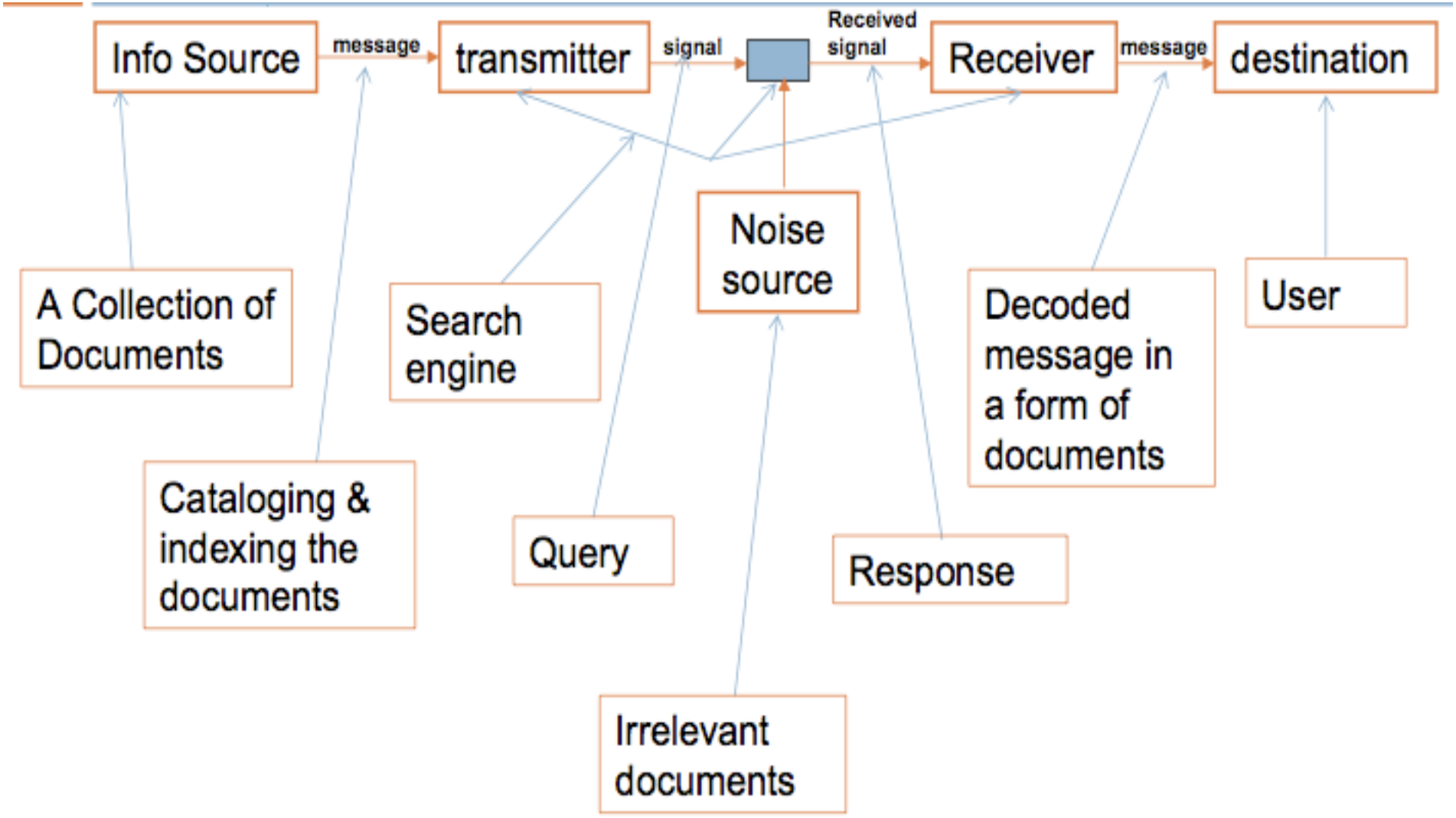
Summary

- Information Theory - A mathematical theory of communication
- Time and cost effectiveness in Signal transmission.
- The design of noiseless channels.
- The semantic aspects of communication are irrelevant.
- Amount of Information -Measures the amount of information that may be generated by an event or a realization of a specific state of affairs.
- The theory tells us how much information a signal carries, but not what information is carried.
- Info sources are defined as sets of events (or states of affairs) with certain probability of occurring.

Summary…

- Semantic Theory of Information, KFI
  - a cognitive, functionalistic theory
  - built on **logical principles** and not on **probability theory**
  - individually oriented, and deals with the content of information
  - Perception, cognition, analogue, digitalization
  - Knowledge, truth, meaning, conceptualization, belief
  - Information as ‘objective commodity’
  - Semantic content, informational content, Semantic content
  - The Xerox principle, characteristics of information, informational content of a signal.

Summary…


Info Source
message
transmitter
signal
Received signal
Receiver
message
destination
A Collection of Documents
Cataloging & indexing the documents
Search engine
Query
Noise source
Irrelevant documents
Response
Decoded message in a form of documents
User