

Introduction to Informatics

Lecture 3: From Information to Informatics



Copyright 1996 Randy Glasbergen. www.glasbergen.com

Readings until now



- Lecture notes

- Posted online @

- <http://informatics.indiana.edu/rocha/i101>

- *The Nature of Information*

- @ *infoport* and web

- *What are blogs?*

- *The Library of Babel* by Jorge Luis Borges

- http://jubal.westnet.com/hyperdiscordia/library_of_babel.html

- From course package

- Von Baeyer, H.C. [2004]. *Information: The New Language of Science*. Harvard University Press.

- Chapters 1, 4 (pages 1-12)

Assignment Situation

- Labs

- Past

- Lab 1: Blogs

- Due this Friday, January 19

- Next

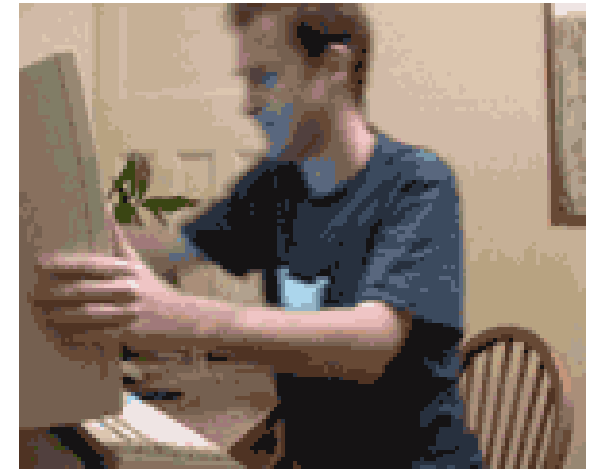
- Lab 2: Basic HTML

- Thursday and Friday, January 18 and 19
 - Due Friday, January 26

- Assignments

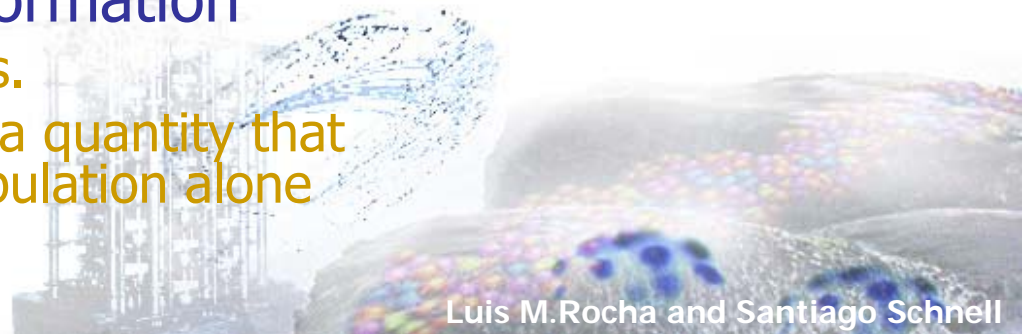
- Individual

- First installment
 - End of January



Information Theory

- article published in 1948 by Claude Shannon
 - “The mathematical theory of communication”
 - Developed to deal with the efficiency of information transmission in electronic channels
- Key concept: information quantity that can be measured unequivocally (*objectively*).
 - Does not deal at all with the *subjective* aspects of information
 - Semantics and pragmatics.
 - Information is defined as a quantity that depends on symbol manipulation alone



What's an information quantity?

- How to quantify a relation?
 - Information is a relation between an agent, a sign and a thing, rather than simply a thing
 - The most palpable element in the information relation is the sign
 - More palpable still is the system of conventional signs we call symbols
- But which symbols do we use to quantify the information contained in messages?
 - Several symbol systems can be used to convey the same message
 - We must agree on the same symbol system for all messages!

Machine

機器

Communication basics

- Both sender and receiver must use the same code, or convention, to encode and decode symbols from and to messages.
 - We need to fix the *language* used for communication
 - Set of symbols allowed (an *alphabet*)
 - the rules to manipulate symbols (*syntax*)
 - the meaning of the symbols (*semantics*).
 - A language specifies the universe of all possible messages
 - Set of all possible symbol *strings* of a given size.
- *Shannon Information* is then defined as “a measure of the freedom from *choice* with which a message is *selected* from the set of all possible messages”

<i>DEAL</i>	EALD	ALDE	LDEA
DELA	EADL	ALED	LDAE
DLEA	ELDA	ADEL	LEDA
DLAE	ELAD	ADLE	LEAD
DAEL	EDLA	AELD	LADE
DALE	EDAL	AEDL	LAED

DEAL is 1 out of 4!
= $4 \times 3 \times 2 \times 1 = 24$
choices.

Information Quantity

- information is defined as the act of selecting a specific message (a string of symbols) from the set of all possible messages (in some language).
- Information content of a message
 - Number of operations needed to select that message from the set of all possible messages
 - Selection process depends on the *likelihood* of occurrence of symbols.
 - depends on the *number of choices* that exist when we encode a message of a given size.
 - Depends on symbols not at all on meaning!
 - "information", "anerthingly" and "innormafito" written in the Roman alphabet with 26 symbols are one of 26^{11} ($=3,670,344,486,987,776 \approx 3.7 \times 10^{15}$) possible words of size 11.
 - Assuming equal likelihood of letters, all have the same information content!
 - Acknowledging different likelihood of letters, which two have the same information content?
 - In the phonetic language of 40 symbols there are 100 times more alternatives!

as $\frac{1}{26^{11}}$ information.

The Library of Babel

Jorge Luis Borges



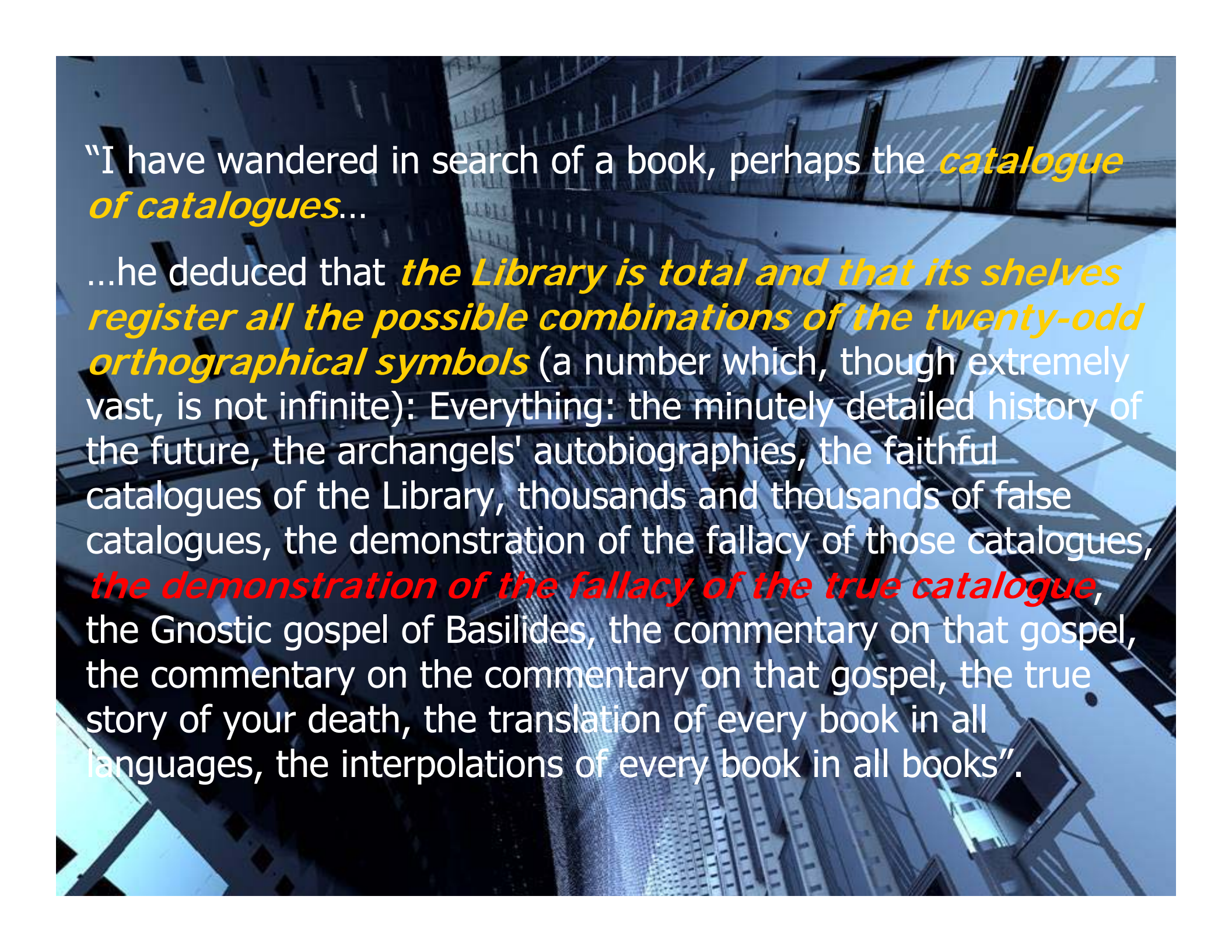
"The universe (which others call the Library) is composed of an indefinite and perhaps infinite number of hexagonal galleries, with vast air shafts between, surrounded by very low railings."

".....all the books, no matter how diverse they might be, are made up of the same elements: the space, the period, the comma, the twenty-two letters of the alphabet. He also alleged a fact which travelers have confirmed: In the vast Library there are no two identical books."

<i>DEAL</i>	EALD	ALDE	LDEA
DELA	EADL	ALED	LDAE
DLEA	ELDA	ADEL	LEDA
DLAE	ELAD	ADLE	LEAD
DAEL	EDLA	AELD	LADE
DALE	EDAL	AEDL	LAED



Luis M.Rocha and Santiago Schnell

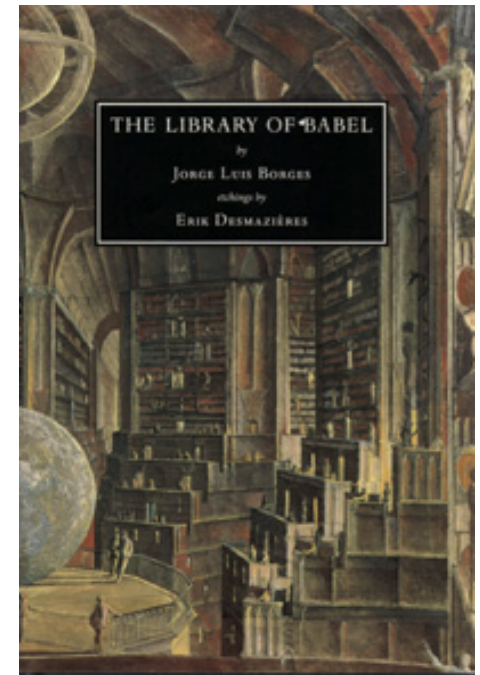


"I have wandered in search of a book, perhaps the *catalogue of catalogues*...

...he deduced that *the Library is total and that its shelves register all the possible combinations of the twenty-odd orthographical symbols* (a number which, though extremely vast, is not infinite): Everything: the minutely detailed history of the future, the archangels' autobiographies, the faithful catalogues of the Library, thousands and thousands of false catalogues, the demonstration of the fallacy of those catalogues, *the demonstration of the fallacy of the true catalogue*, the Gnostic gospel of Basilides, the commentary on that gospel, the commentary on the commentary on that gospel, the true story of your death, the translation of every book in all languages, the interpolations of every book in all books".

Numbers in the Library of Babel

- 25 characters, written in any sequence for 410 pages of 40 lines of 80 characters = $25^{(410 \times 40 \times 80)}$ combinations = 251,312,000 books!
 - = **101,834,097** books
- Total number of atoms in the current, observable universe is about 10^{80}
- So if each book was the size of an atom, the library would hold $10^{1,834,017}$ **universes** worth of books!
- Throw away the equivalent of a single grain of sand—100,000,000,000,000,000,000 books—and, though an enormous number, it would be as nothing



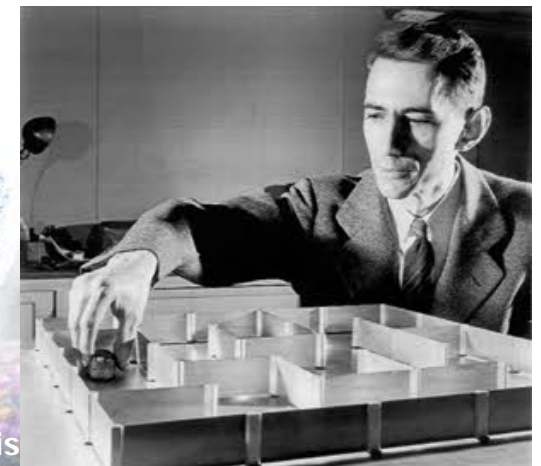
“One: the Library is so enormous that any reduction of human origin is infinitesimal. The other: every copy is unique, irreplaceable, but (since the Library is total) there are always several hundred thousand imperfect facsimiles: works which differ only in a letter or a comma.”

Information Theory

- Claude Shannon's formulation
 - Efficiency of information transmission in electronic channels
 - Key concept
 - *information quantity* that can be measured unequivocally (*objectively*).
 - Does not deal with *subjective* aspects of information
 - Language must be agreed (fixed) by sender and receiver
 - Same convention, or *encoding*
 - Symbols allowed (an *alphabet*) and syntax (rules) and *semantics*.
 - Specifies the universe of all possible messages
- Shannon Information is then defined as "a measure of the freedom from choice with which a message is *selected* from the set of all possible messages"

<i>DEAL</i>	EALD	ALDE	LDEA
DELA	EADL	ALED	LDAE
DLEA	ELDA	ADEL	LEDA
DLAE	ELAD	ADLE	LEAD
DAEL	EDLA	AELD	LADE
DALE	EDAL	AEDL	LAED

DEAL is 1 out of 4!
= $4 \times 3 \times 2 \times 1 = 24$
choices.



Binary Code

- Language with an alphabet of two symbols
 - "0" and "1", "FALSE" or "TRUE", etc.
- Most economical way of encoding information
 - Example from von Baeyer's [2004] book *"Information: The new language of Science"*
 - Consider a sailor who wants to signal a number between 0 and 127 by means of flags.

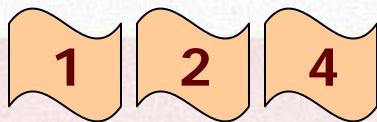
0: 0	4: 100
1: 1	5: 101
2: 10	6: 110
3: 11	7: 111

Message from Ship

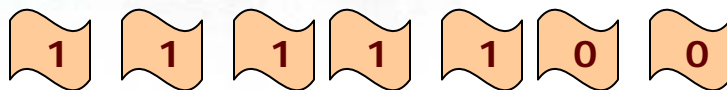
System 1: one flag per number



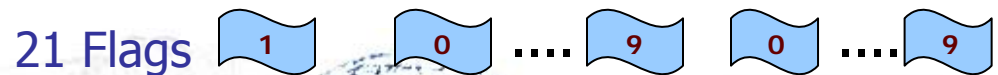
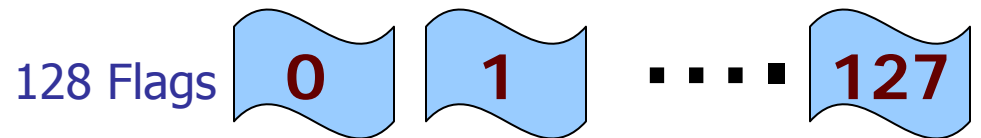
System 2: decimal system



System 3: binary system



Required Flags



The Bit

- Shannon used the binary system because it is the most economical
 - Uses less memory
 - Information quantity depends on the number of alternative message *choices* encoded in the binary system
- Bit (short for *binary digit*) is the most elementary choice one can make
 - Between two items: "0" and "1", "heads" or "tails", "true" or "false", etc.
 - Bit is equivalent to the choice between two equally likely choices
 - Example, if we know that a coin is to be tossed, but are unable to see it as it falls, a message telling whether the coin came up heads or tails gives us one bit of information

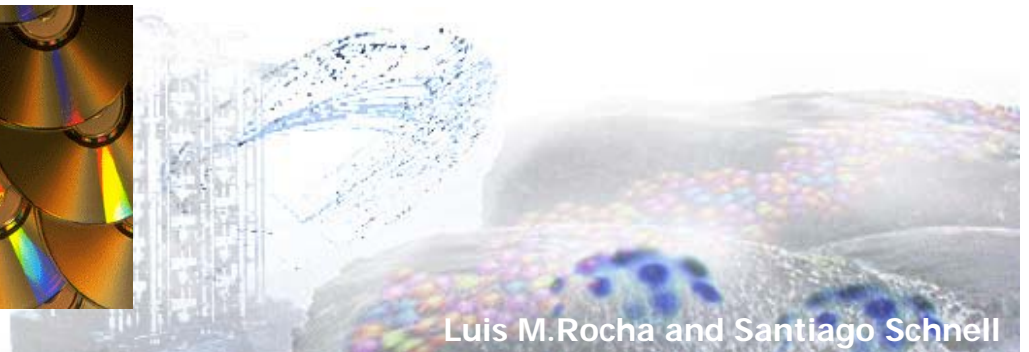
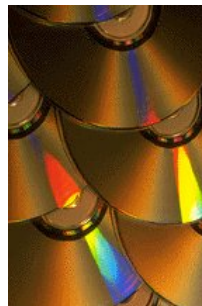
Digital versus Analog



- *Digital* is used to convey the notion of discrete objects/values
 - Things we can count
 - The word digit comes from the Latin word for finger (*digitus*)
 - Digital information is equivalent to symbolic information
 - Any symbol system requires a set of discrete symbols for setting up an arbitrary semantic relation



- **Analog (or Analogue)**
 - Information transmission via electrical, mechanical, hydraulic, and sound signals
 - Continuously varying signals which are not countable
 - What was used up until Shannon
 - Instead of messages being arbitrarily encoded, analog signals rely on some physical property of the medium
 - It implies an analogy between cause and effect, input and output
 - Voltage as an "analogy" to sound in analog synthesizer
 - But it cannot encode any sound whatsoever!
 - Sounds depend on the physical properties of electricity, the transducer and equipment used



Analog Synthesizers

- Are not “universal” sound devices
 - Like sample synths and digital sound cards
 - Their sound depends very much on the analogs employed
- Famous synths
 - Minimoog
 - <http://www.synthmuseum.com/moog/moomini01.html>
 - Roland TB-303



Semiotics and Informatics

Information

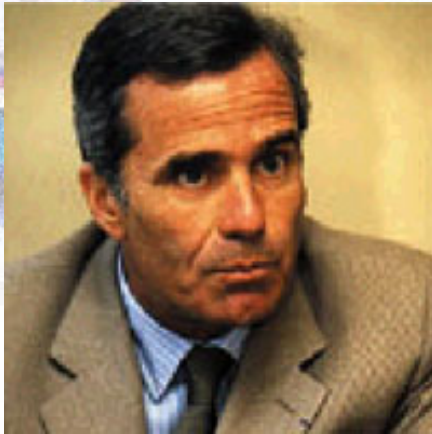
- **Semantics**
 - the content or **meaning** of the Sign of a Thing for an Agent
 - Relations between signs and objects for an agent
 - the study of meaning.
- **Syntax** \leftarrow **Information Technology**
 - the characteristics of signs and symbols devoid of meaning
 - Relations among signs such as their rules of operation, production, storage, and manipulation.
- **Pragmatics**
 - the context of signs and repercussions of sign-systems in an environment
 - it studies how context influences the interpretation of signs and how well a signs-system represents some aspect of the environment

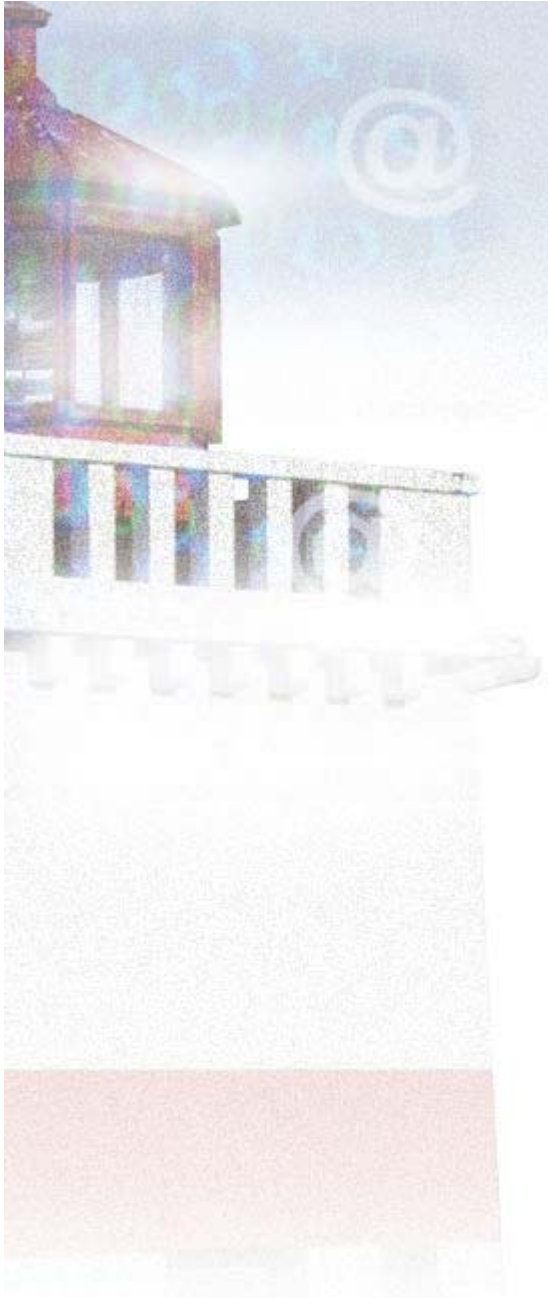
Informatics

Origins of "Informatics"

The French Connection

In 1962 in France, a group started a software company called Societe pour L'Informatique et Applique (SIA). Philippe Dreyfus, a French information system/software pioneer, was a principal founder. Dreyfus coined the word as a combination of "information" and "automatic."





SIA - Informatics

- Came from applied information science
- Concerned with sci/tech communication
 - communication processes in this community
 - development of more efficient systems/techniques



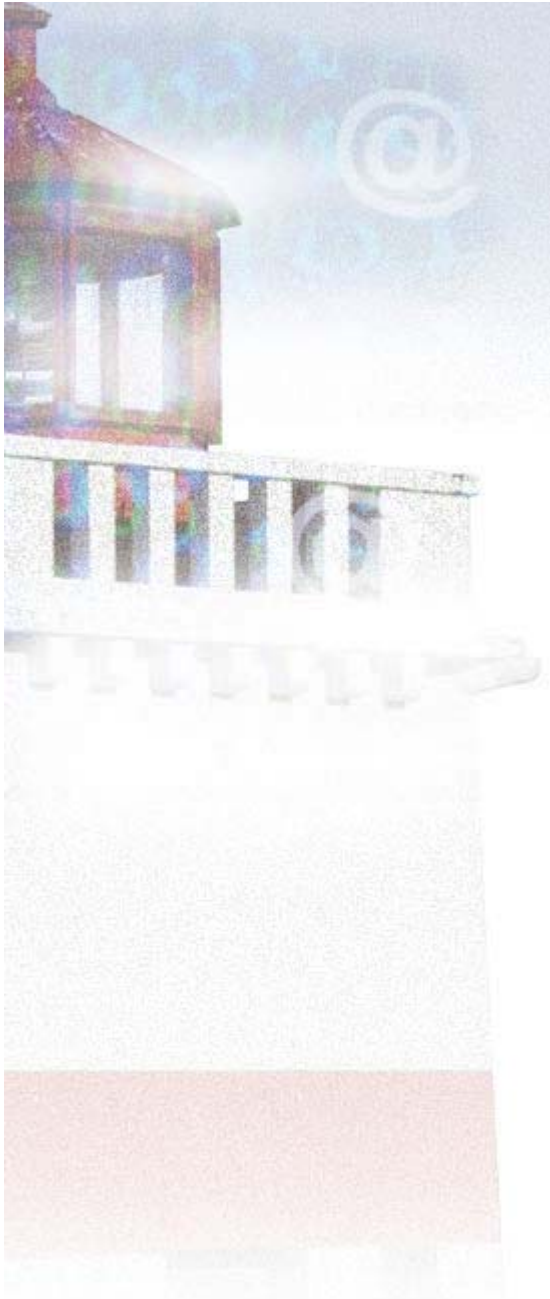


Origins of “Informatics”

The American Connection

Also in 1962 in the U.S., Walter F. Bauer founded a company named Informatics. Later the Association for Computing Machinery officially applied for permission to use the name. They wished to change ACM to Society for Informatics. But they were turned down on the advice of Bauer’s lawyers.

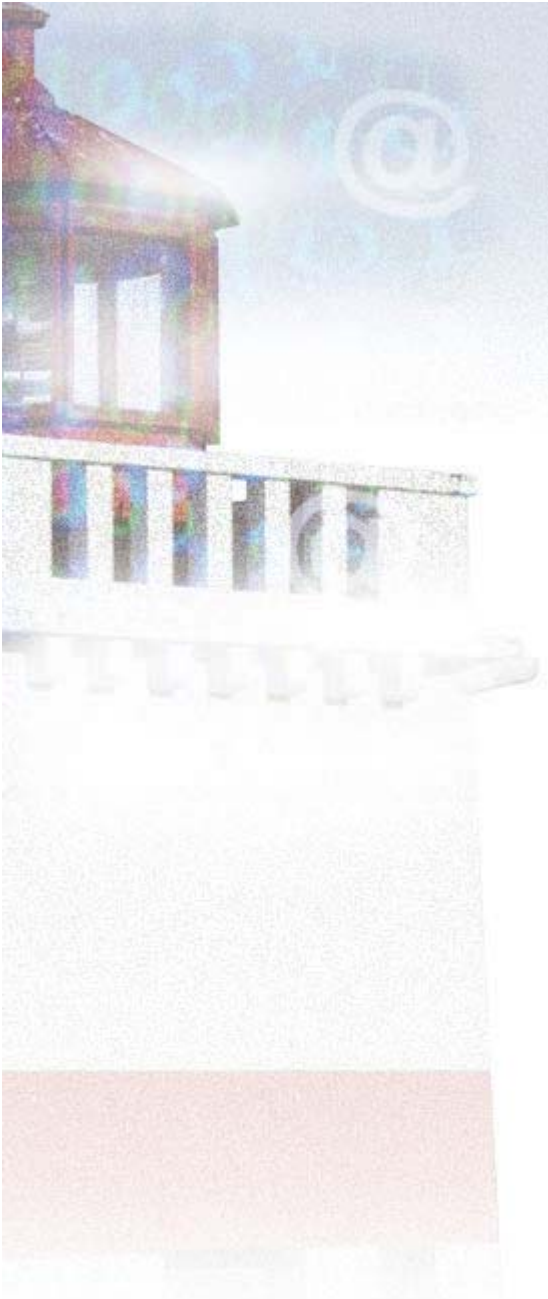




Historical usage of Informatics

- Popularized by Soviets (Informatika)
 - Considered a branch of social sciences
- Other countries (including France)
 - Considered it to be applied computer science
- In the U.S.
 - Continued to use term "computer science"



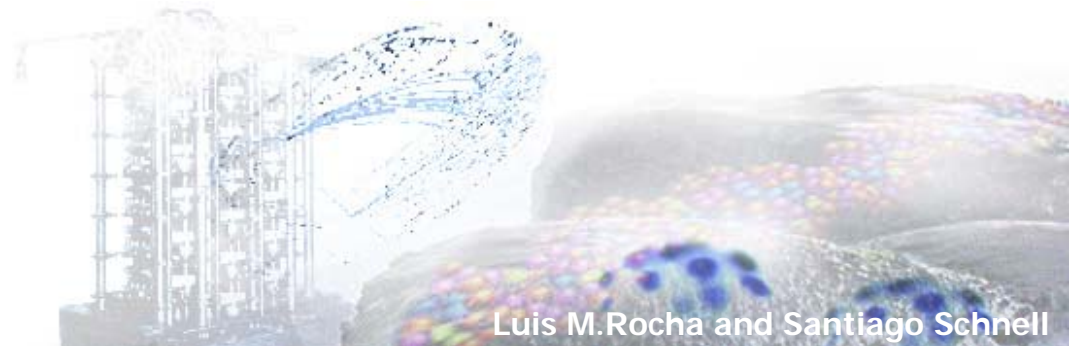


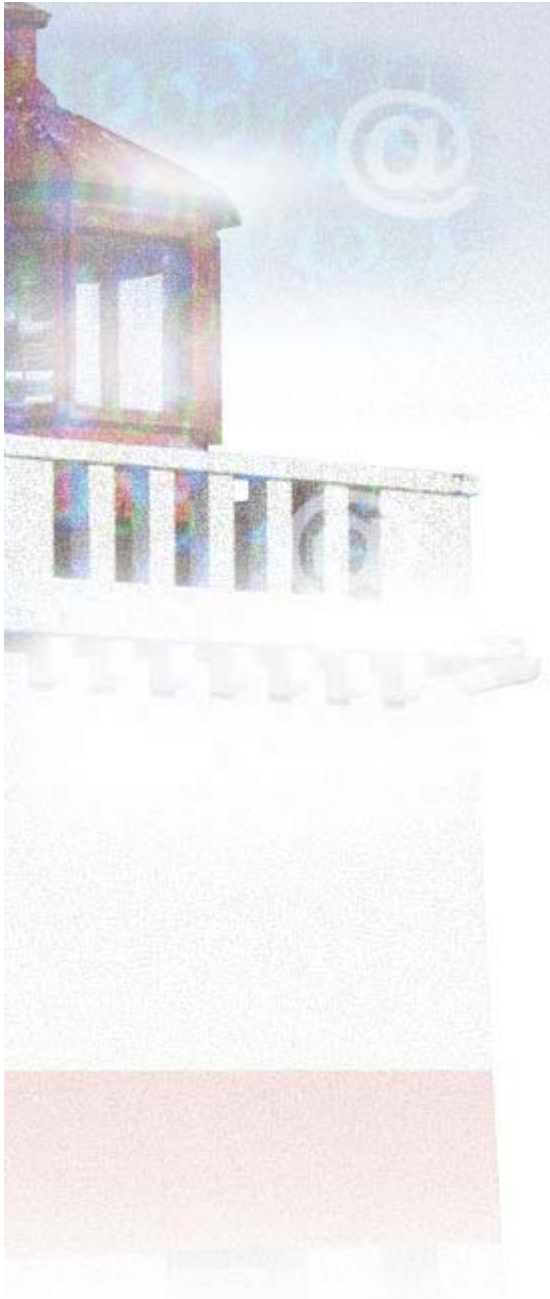
Usage of Informatics

In the USA

“During the late 1970’s, several fields of Information technology application flirted with using the term, but of these only one – Medical Informatics - caught hold and remained.”

UCI Informatics Task Force Report 1996



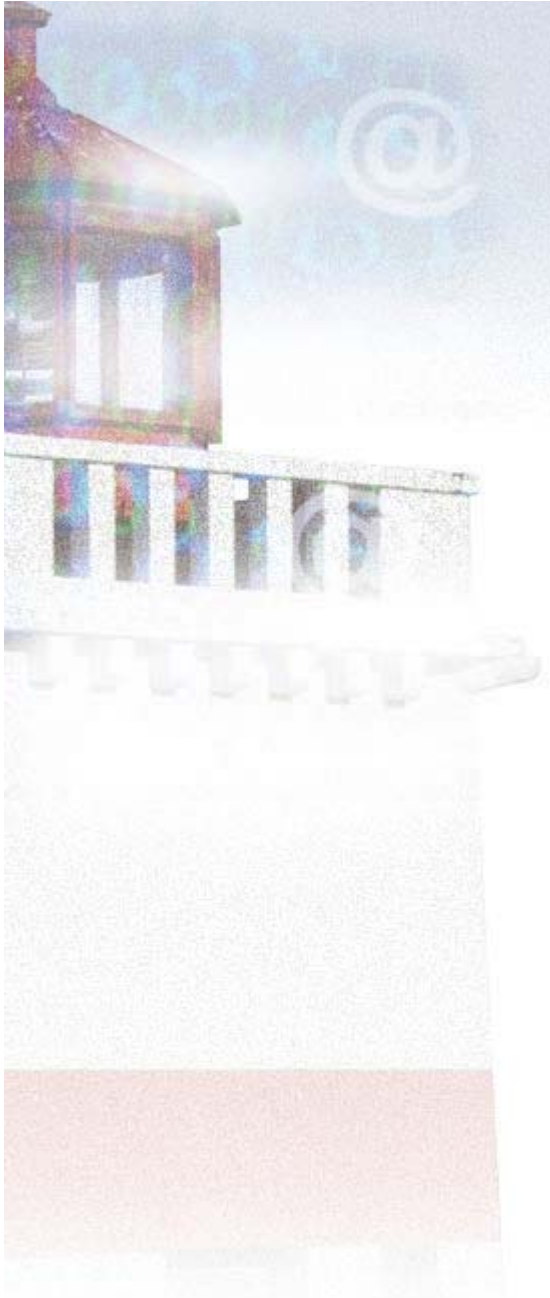


Informatics in the dictionary

The French Connection

The word "informatique" took on the meaning, generically, of "electronic information processing." It was accepted (1966) as an official French word by L'Academie Francaise, something not easy to do. The word "informatique" has now been adopted and adapted in Europe to mean roughly "computer science."





Informatics in the dictionary

The American Connection

The word “informatics” has been widely used in application contexts: medical informatics, nursing informatics, legal informatics, chemical informatics, etc., and of course the currently hot bioinformatics.

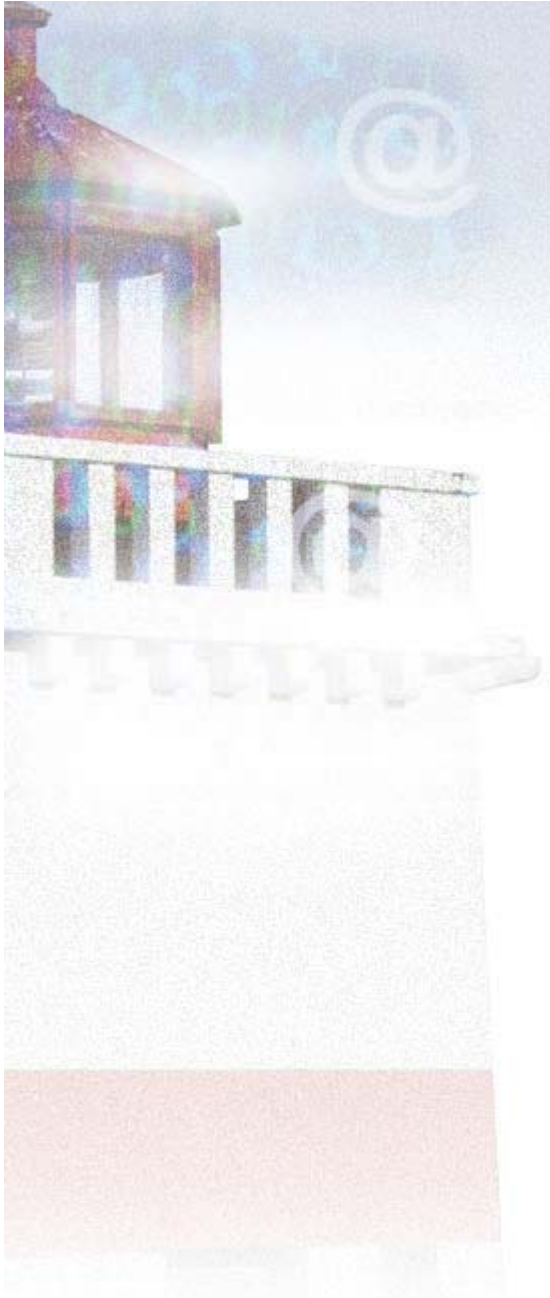




Informatics – New name of Theory of Scientific Information

According to Mikhailov et al (1967), **Informatics** is the discipline of science which investigates the structure and properties (not specific content) of scientific information, as well as the regularities of scientific information activity, its theory, history, methodology and organization. The purpose of informatics consists in developing optimal methods and means of presentation (recording), collection, analytical-synthetic processing, storage, retrieval and dissemination of scientific information. Informatics deals with logical (semantic) information, but is not involved in qualitative estimation of this information. Such an estimate can be carried on by specialists alone, in the specific fields of science or practical activity.

Informatics = information + the "tic" which in Greek means theory (arithmet*ic*, aeronaut*ics*, etc.)



Informatics at IU

At Indiana University we have built upon this usage of "x-informatics" or "hyphen-informatics" to establish (1999) the School of Informatics, which focuses on domain specific applications of information technology, as well as upon the social and organizational issues associated with its use (social informatics, organizational informatics). We also focus on the commonalities of applications (largely driven by large data sets), including interfaces, visualization, digital media, data mining, complex systems, simulation, modeling, etc. We include logic (and probability, statistics), language (natural language interfaces), cybersecurity, and information theory.

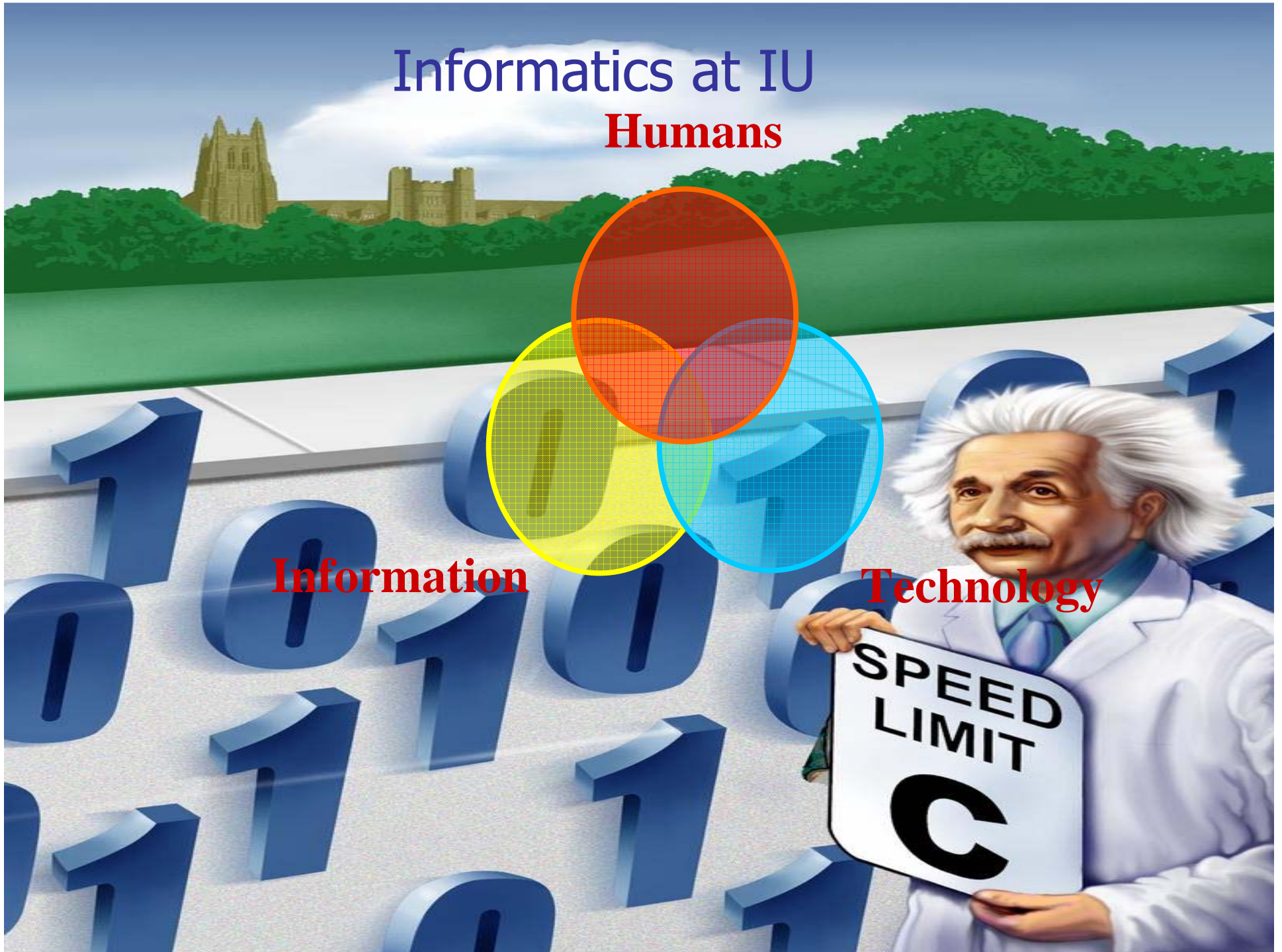


Informatics at IU

Humans

Information

Technology





Next Classes!

■ Topics

- Digital vs. Analog
- What is Technology?
 - What is Information Technology?
 - Examples of important IT

■ Readings for Next week

- Lecture notes Posted online @ <http://informatics.indiana.edu/rocha/i101>
 - *Technology*
- @ *infoport*
- From course package
 - From Andy Clark's book "*Natural-Born Cyborgs*"
 - Chapter 2: "Technologies to Bond With" (pages 19 - 44)

■ Lab

- Basic HTML