Information or What is stuff?

CS 431 Architecture of Web Information Systems

Carl Lagoze
Cornell University – Spring 2008

Logistics

- Start thinking about reaction papers
- Partners for projects chosen by Feb 15
- Posts should be made by 8AM Friday AM
- Don't wait for the end of the semester to establish a blog presence
- Connection of readings to lecture and intelligent questions

Acknowledgments

- Herbert Van de Sompel – Los Alamos Research Library
- Barbara Tillet - Chief, Cataloging Policy and Support Office Library of Congress
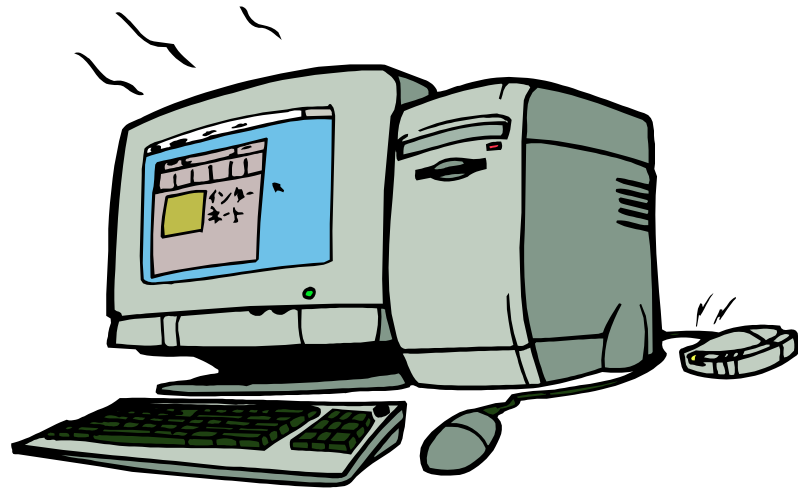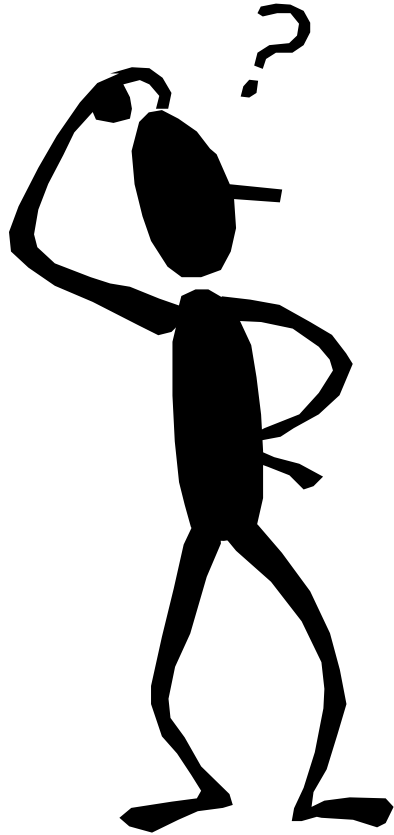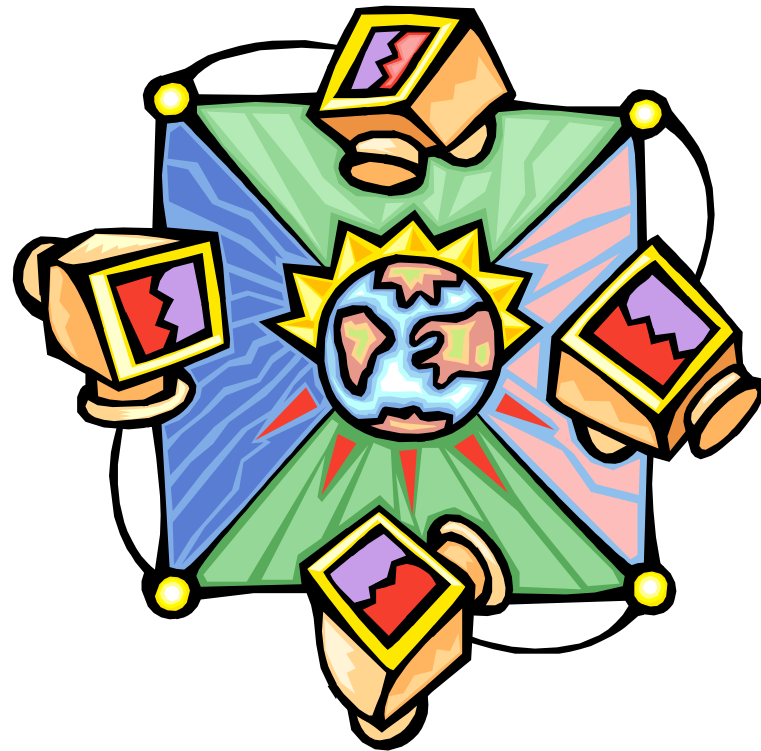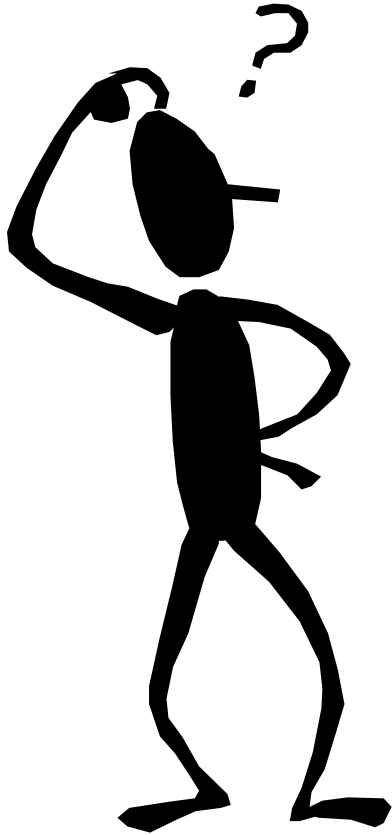
People want stuff.

Godfrey Rust 1999

Stuff is more complex than you think.

Carl Lagoze 2008

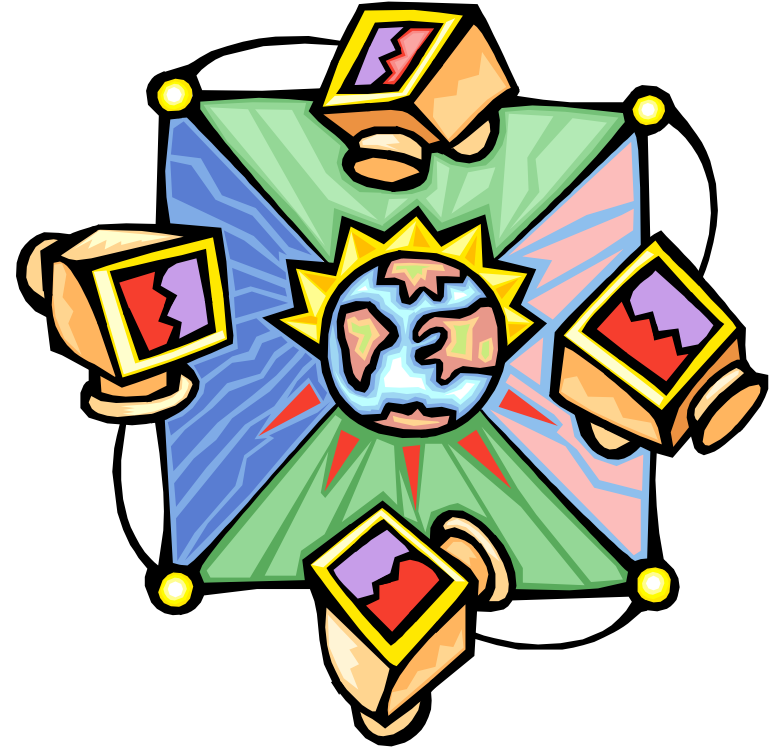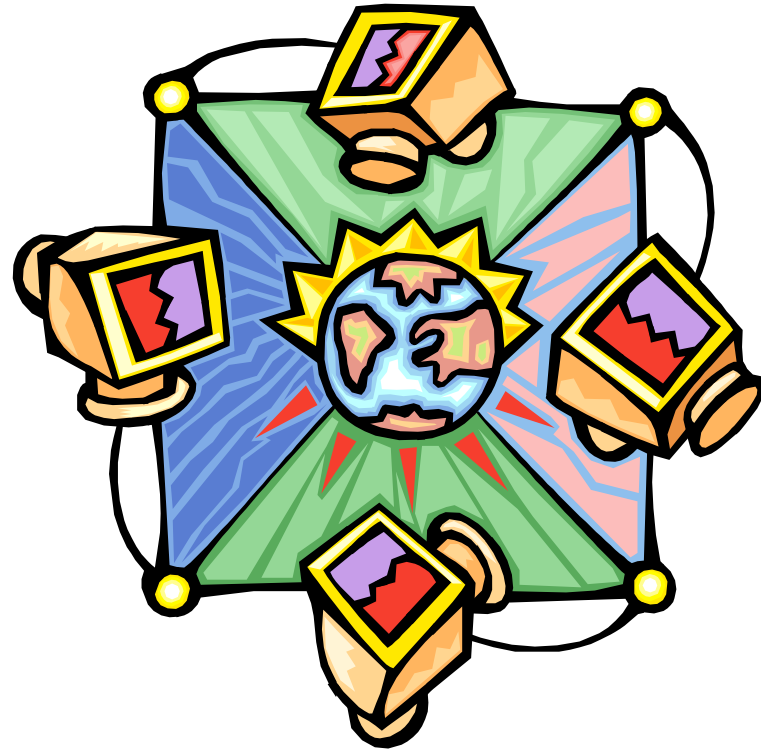Where did I put that file?

Where is that information?

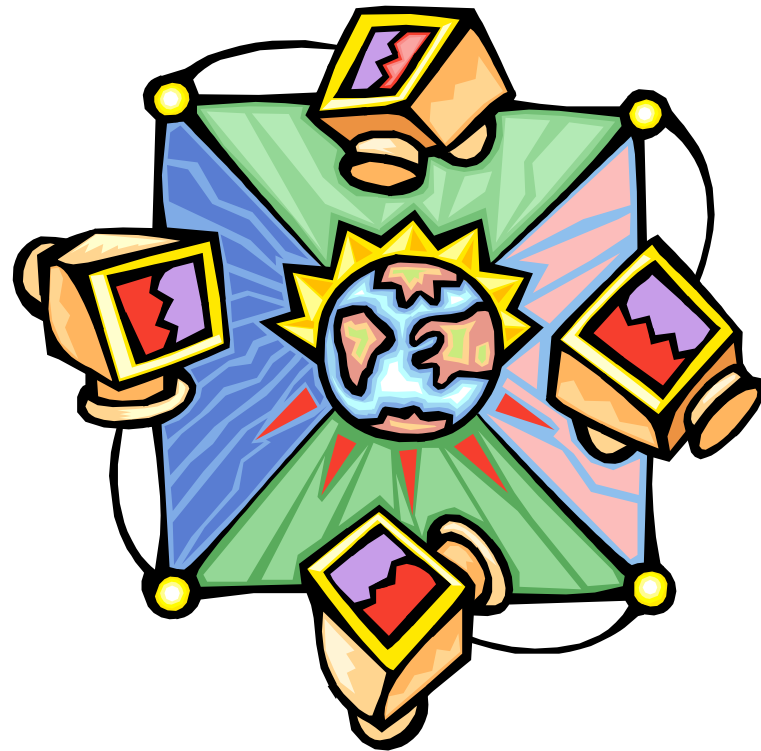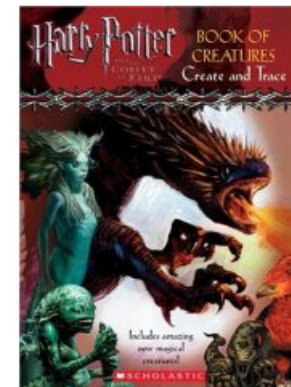# Am I getting compensated for my talent?
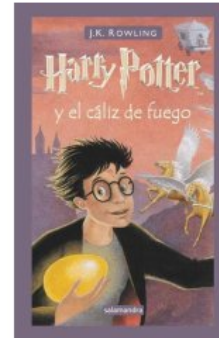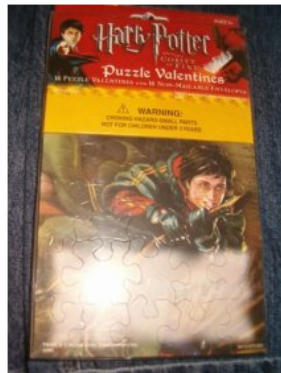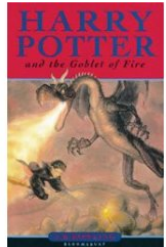

Copies?

Derivations?

Contributions?

Is that available in a way I can use it?

Are there other resources like these?

So you said you wanted "Harry Potter and the Goblet of Fire"
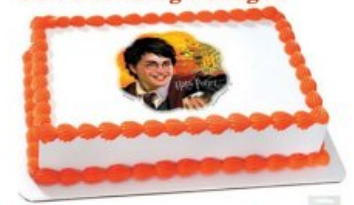
Or you want to read "As we may think"

Google Scholar BETA

vannevar bush as we may think | Search

As **we may think** - Get it! Cornell - group of 66 »
V **Bush** - interactions, 1996 - portal.acm.org
Page 1. 35 interactions . . . march 1 9 9 6 As **We May Think Vannevar Bush** As Director of the Office of Scientific Research and Development, ...
Cited by 1887 - Related Articles - Web Search - BL Direct

[CITATION] As **we may think**
B **Vannevar** - The Atlantic Monthly, 1945
Cited by 8 - Related Articles - Web Search - Get it! Cornell

[BOOK] From Memex to hypertext: **Vannevar Bush** and the mind's machine
JM Nyce, P Kahn - 1991 - Academic Press Professional, Inc. San Diego, CA, USA
Cited by 65 - Related Articles - Web Search

[CITATION] The Atlantic Monthly
V **Bush**, AWM **Think** - Vol. t76, 1945
Cited by 11 - Related Articles - Web Search

[CITATION] As **we may think**, interactions, v. 3 n. 2
V **Bush** - 1996 - March
Cited by 6 - Related Articles - Web Search

As we may think - Get it! Cornell
V Bush - interactions, 1996 - portal.acm.org
Page 1. 35 interactions . . . march 1 9 9 6 As We May Think Vannevar Bush As Director of the Office of Scientific Research and Development, ...
Cited by 1887 - Related Articles - Web Search - BL Direct

As We May Think
V Bush - Atlantic Monthly, 1945 - makebelieve.gr
Page 1. The Atlantic Online | July 1945 | As We May Think | Vannevar Bush http://www.theatlantic.com/doc/print/194507/bush 1 of 8 21/3/2006 8:46 πμ ...
View as HTML - Web Search - Get it! Cornell

As we may think
V BUSH - Library computing(Thousand Oaks, Calif.), 1999 - cat.inist.fr
As we may think. Vannevar BUSH Library computing(Thousand Oaks, Calif.) 18:33, 180-188, Sage, 1999. L'auteur incite les hommes de ...
Web Search

" As We May Think"--
V Bush, O Thursday, O Friday - eecs.mit.edu
MIT Department of Electrical Engineering & Computer Science. "As We May Think" --. A Celebration of Vannevar Bush's 1945 Vision, An ...
Cached - Web Search

As We May Think
V Bush - Atlantic Monthly, 1945 - mirror.eacoss.org
Page 1. The Atlantic Monthly | July 1945 As We May Think BY V ANNEVAR B USH ..... As Director of the Office of Scientific Research ...
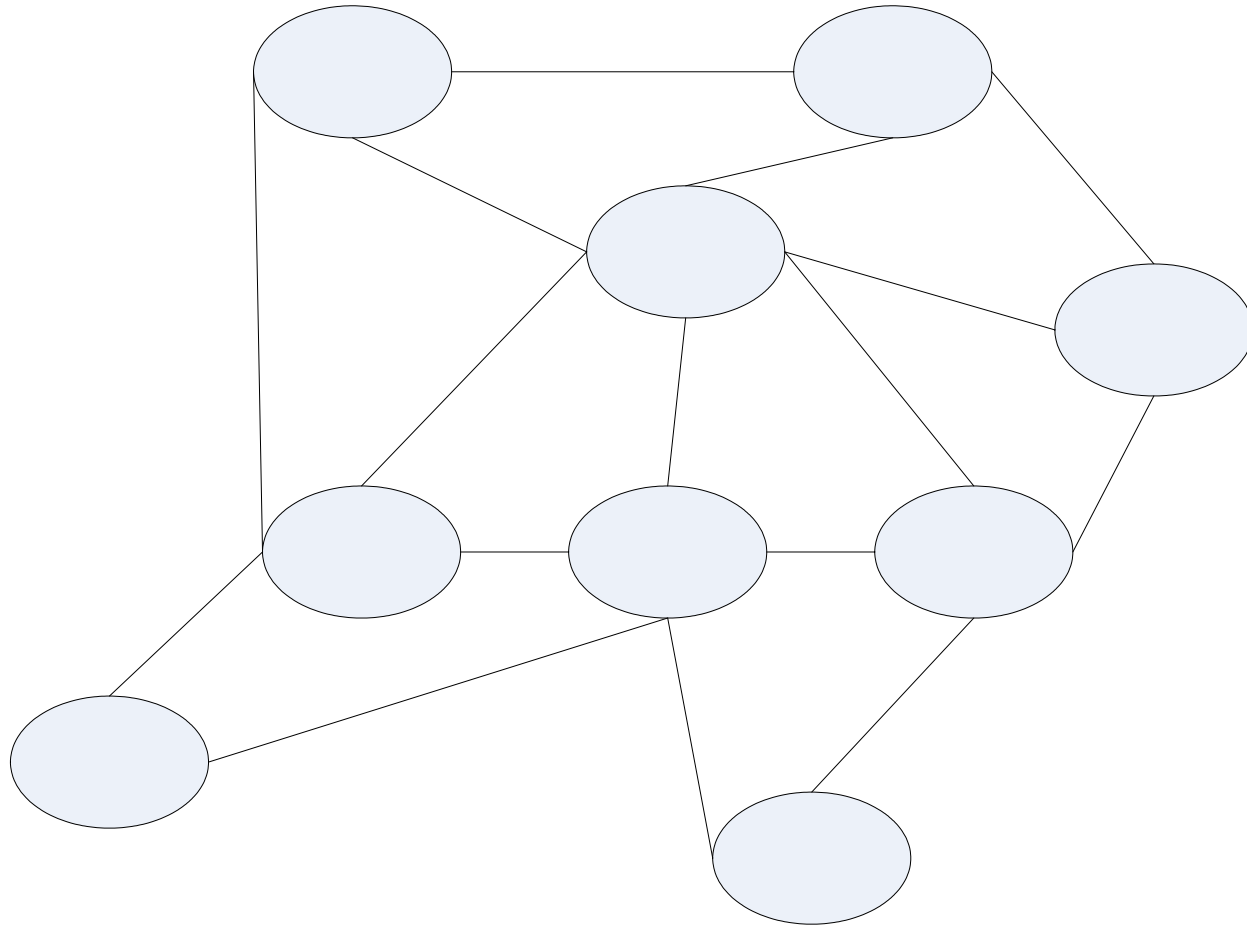View as HTML - Web Search - Get it! Cornell

As We May Think
V Bush - ariel.adgrp.com
Next: Introduction As We May Think. by Vannevar Bush. This article was originally published in the July 1945 issue of The Atlantic Monthly . ...
Cached - Web Search

From individual web graph nodes to compound documents

From individual web graph nodes to compound documents
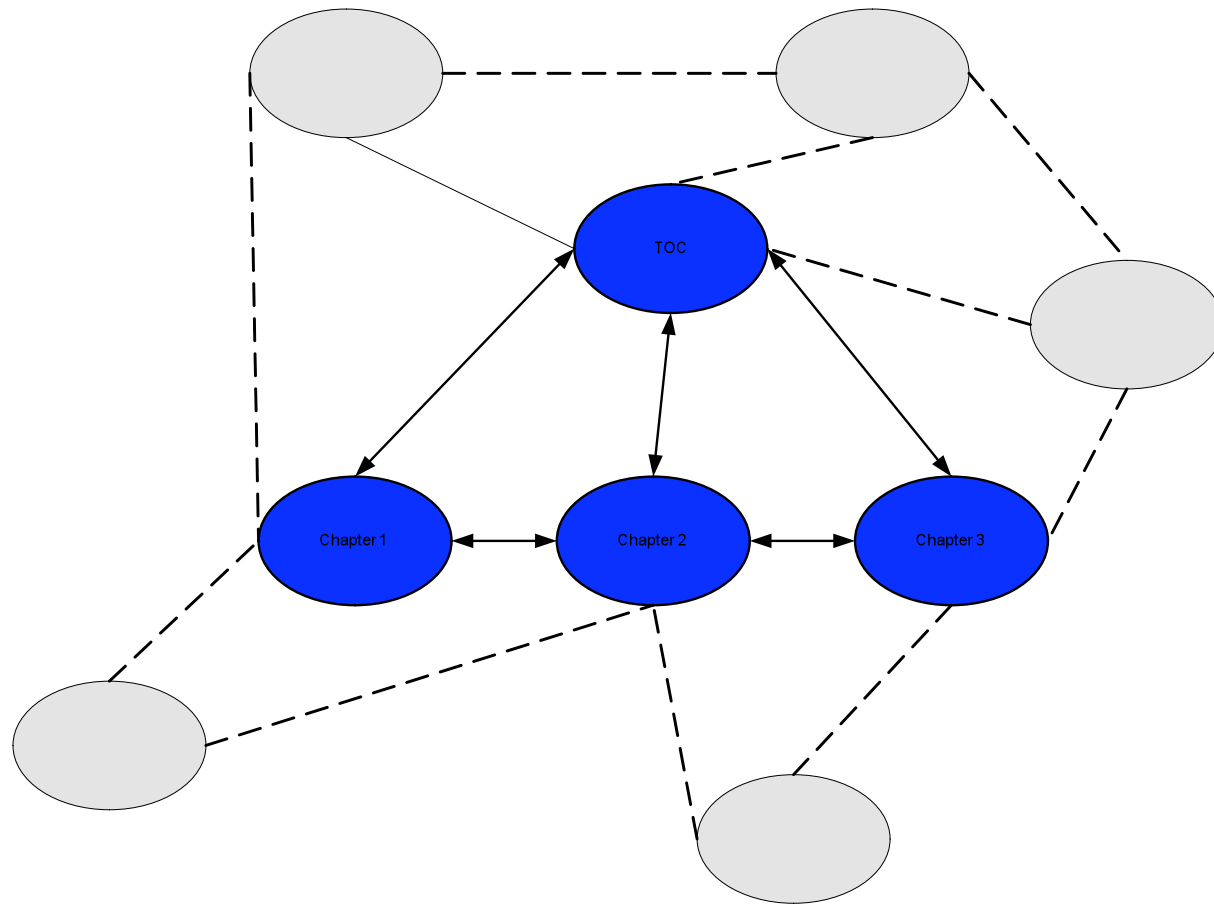
Information vs. Data – Formal basis

- Claude Shannon – Problem of noisy communication channels
- **Entropy**
  - Informally - measure of the amount of information in a data transmission
  - Amount of disorder in a system.
  - Proportional to the uncertainty of the recipient of a data stream about the content of the message

- Implications
  - Same content can be **encoded** in multiple data streams
  - These different streams mathematically have different information content even though they may be expressions of same information (preservation implications)

Same Information?

24-bit color

8-bit color

Content, Data, Metadata Surrogates -- **informal definitions**

Content refers to resources as **information** that is of interest to a user. It is the human view of information:

music	**Beethoven's Fifth Symphony**
database	**Genome Database**
literature	**Gone with the wind**
web site	**weather.com**
software	**MS Word**

Content, Data, Metadata Surrogates -- **informal definitions**

Data emphasizes on the bits and bytes to be processed by a computer. It is the computer representation of information:

bit and byte layout (e.g., ASCII)
compression schemes (e.g., MP3)
image format (GIF, JPEG, PNG)

Content, Data, Metadata Surrogates -- **informal definitions**

Metadata Surrogates is data about data/content.

      Descriptive metadata (e.g., catalog records)

      Administrative metadata (e.g. circulation records)

      Structural metadata (e.g. serials record)

      Rights metadata (e.g. shrink wrap license)

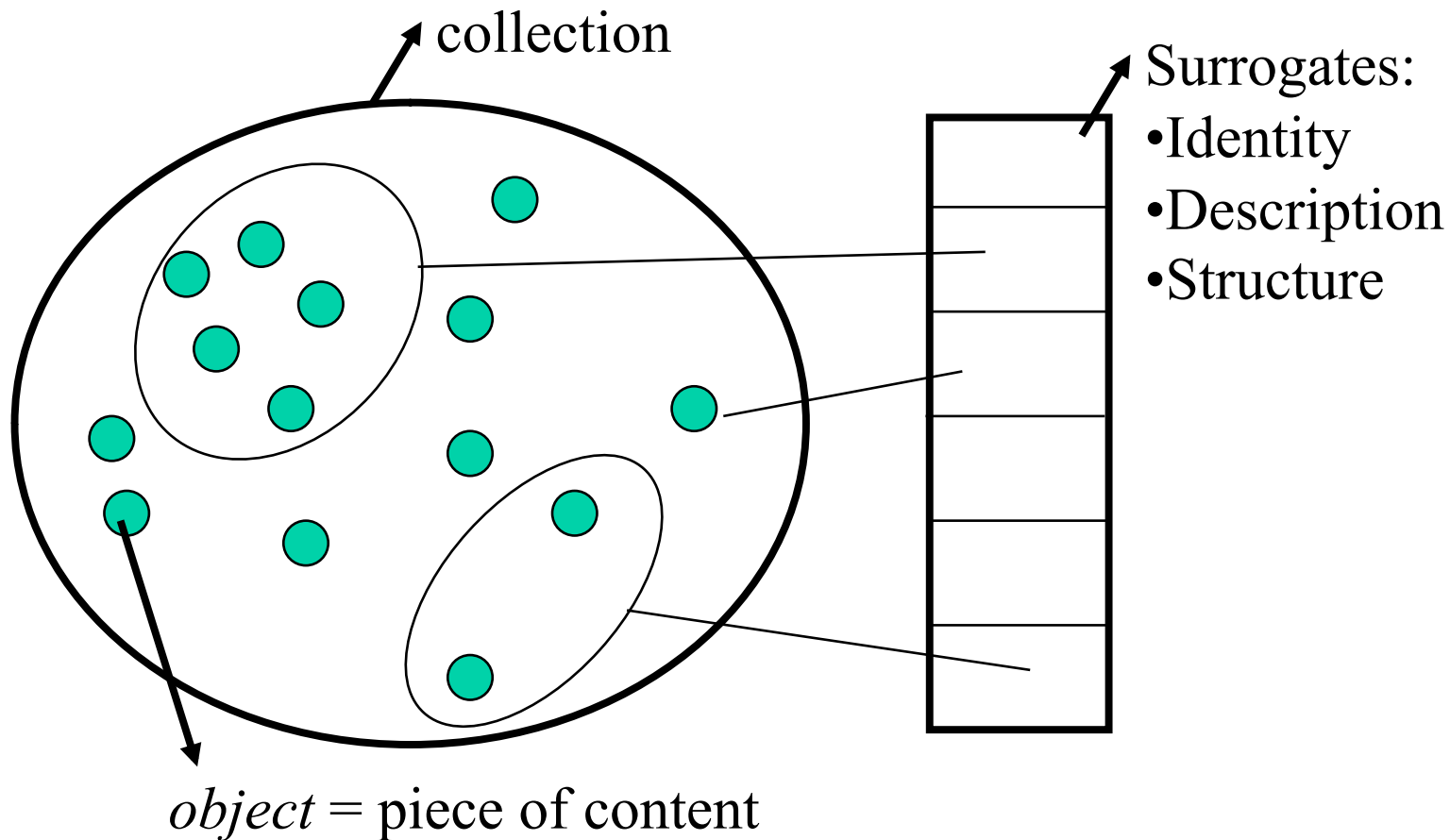Provides a tangible **surrogate** for:

      Abstract information entities

      Available or unavailable physical information entities

Information (Bibliographic) model

establishes equivalence classes to organize information objects for human understanding and management

collection

Surrogates:
• Identity
• Description
• Structure

object = piece of content

Objectives of a bibliographic system

1. To locate **objects** in a file or database as the result of a search using attributes or relationships of the **objects**:
- Known item search -To find a singular object
- Search - To locate sets of objects representing:
  - All **objects** corresponding to some information unit
  - All **objects** by a given author
  - All **objects** about a given author
  - All **objects** on a given subject
  - All objects published by a given publisher
- These are all access points into the information space

Objectives of a bibliographic system

2. To identify an **object** (i.e. confirm that a described object corresponds to the sought object or distinguish between objects with similar characteristics)

3. To select an **object** that is appropriate to the user's needs/capabilities

4. To obtain access to an **object** (purchase, loan, license, …)

5. To navigate the file or database (browse, serendipitous discovery)

Traditional models challenged by networked digital information

- Scale of corpus or collection
- Variety of content – **Internet Commons**
- Unbinding of information from its carrier
- Mutability/dynamic nature of content
- 'universal context' – all types of people, resources, needs

Requires more advanced data models to represent:
- Distinct entities
- Their attributes
- Their relationships
- Their evolution over time

Entity-Relationship Model

- ## Entities
- ## Relationships
- ## Attributes

"attr"

| Entity 1 | relationship | Entity 2 |
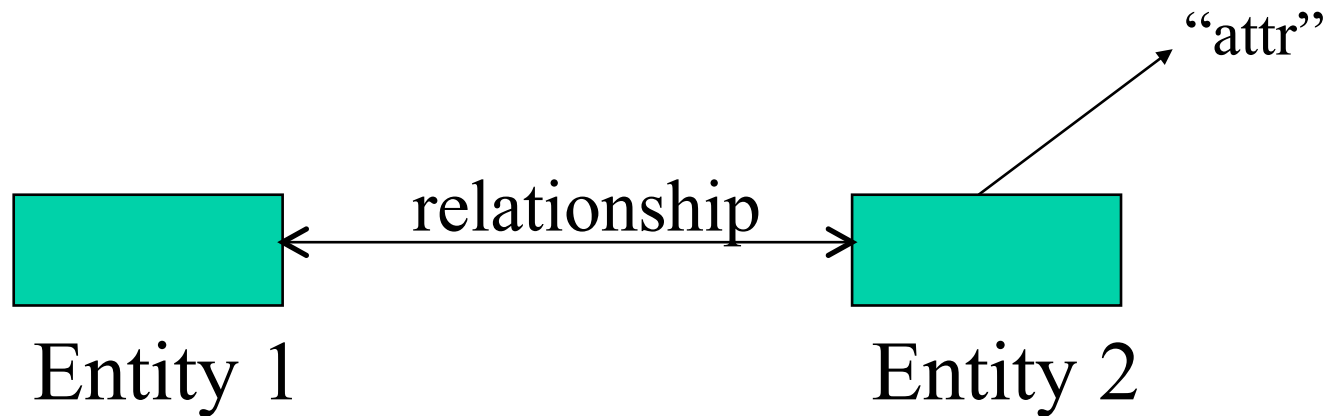
We'll see this applied later in the semester in the semantic web

Variants of information entities: Classes with different attributes

- Psycho Killer

- The score by David Byrne
- The song as played by the Talking Heads

- Live "Stop Making Sense" performance in 1993 in Hollywood
- DVD of 1993 Performance
- The 45 RPM single released in 1979

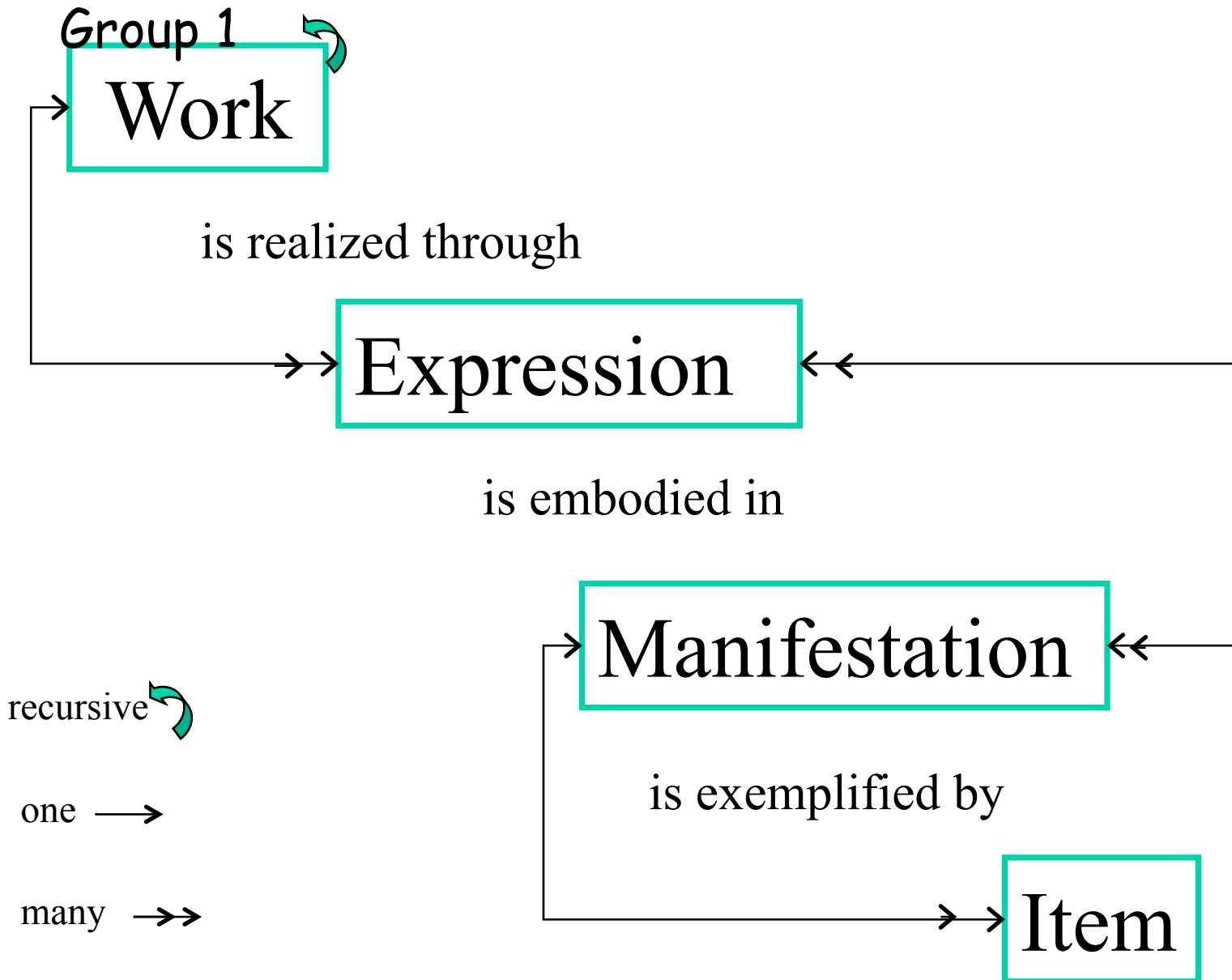- Carl's personal copy of the DVD "Stop Making Sense"

IFLA Model to represent object variants: entities

Entities are the key objects of interest to users of bibliographic data (i.e. of a bibliographic system):

Group 1 - products of intellectual endeavor: work, expression, manifestation, item

Group 2 – the parties responsible for the intellectual content: person, corporate body

Group 3 – the subjects of intellectual endeavor: concept, object, event, place

Group 1

Work

is realized through

Expression

is embodied in

Manifestation

is exemplified by

Item

recursive

one

many

IFLA Model: work, expression, manifestation, item

A work is an **abstract** entity, an idealization e.g.

- The Iliad

- The Weather Channel web site

- Beethoven's Fifth Symphony

- Unix operating system

- The Bible

This is roughly equivalent to the concept of "literary work" used in copyright law.

IFLA Model: work, expression, manifestation, item

An expression is a realization of a work; a representation of the work in a disseminatable form e.g.

- **The Iliad** has oral expressions and written expressions

- A musical work has a score, live performance(s), an original recording, cover(s), ...

Many works have only a single expression, e.g. a web page, or a book which only has a single edition, a painting, a medieval manuscript.

Modes of Expression

- Alpha-numeric
- Music notation
- Other notation (e.g., choreographic)
- Sound
- Still image
- Moving image
- Three-dimensional object
- Combinations of above

IFLA Model: work, expression, manifestation, item

A manifestation is the concrete embodiment of an expression; it reflects <u>physical</u> form e.g.

- The text of **The Iliad** has been manifested in numerous manuscripts and printed books.

- A musical recording can be distributed on CD, cassette, or on a soundtrack of a DVD.

IFLA Model: work, expression, manifestation, item

When many copies are made of a manifestation, each copy is a separate item, e.g.

- the Cornell Library's copy of an edition of the Iliad
- the mp3 of the latest Radiohead CD on Carl's iPod

Group 1 Entities' Attributes

FRBR

- **Work**
  - ID
  - Title
  - Date
  - etc.

- **Expression**
  - ID
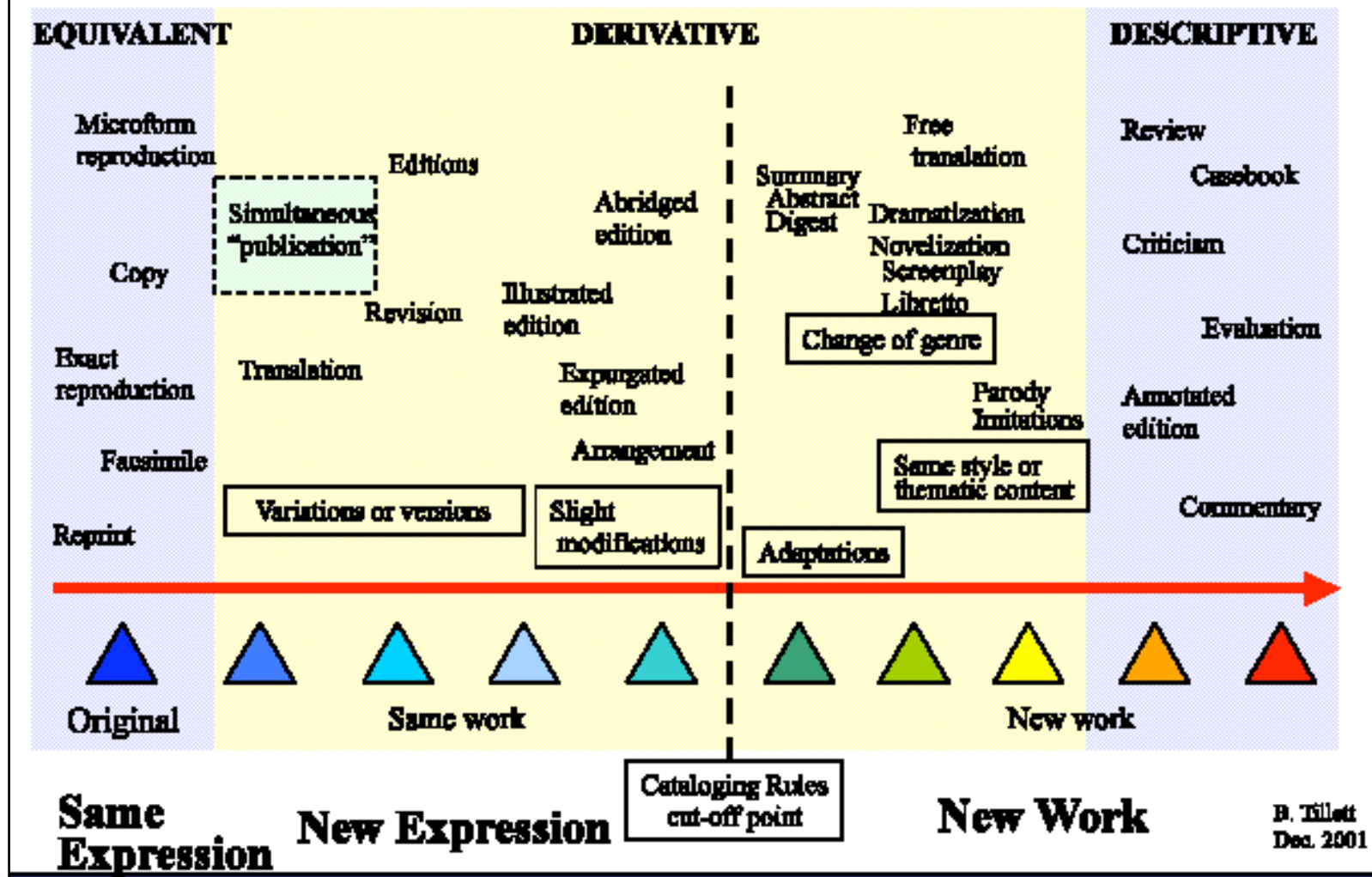 - Title
  - Form
  - Date
  - Language
  - etc.

- **Manifestation**
  - ID
  - Title
  - Statement of responsibility
  - Edition
  - Imprint (place, publisher, date)
  - Form/extent of carrier
  - Terms of availability
  - Mode of access
  - etc.

- **Item**
  - ID
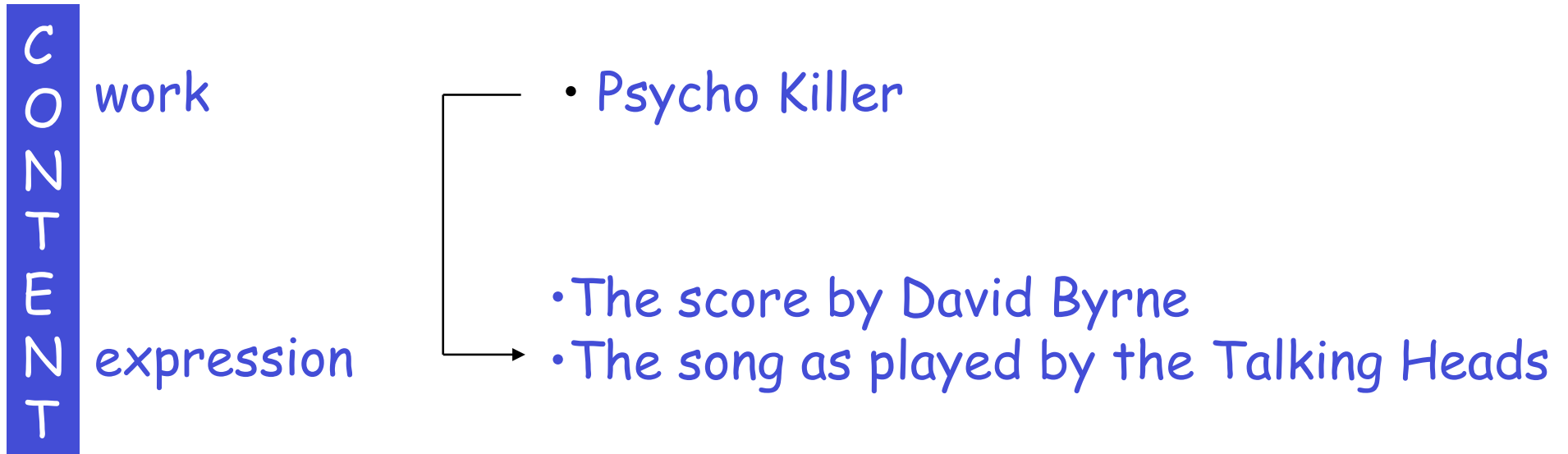 - Provenance
  - Location
  - etc.

Family of Works

EQUIVALENT   **DERIVATIVE**   **DESCRIPTIVE**

Microform reproduction

Editions

Free translation

Review

Casebook

Simultaneous "publication"

Abridged edition

Summary Abstract Digest

Dramatization

Criticism

Copy

Revision

Illustrated edition

Novelization Screenplay Libretto

Change of genre

Evaluation

Exact reproduction

Translation

Expurgated edition

Parody Imitations

Annotated edition

Facsimile

Arrangement

Same style or thematic content

Commentary

Reprint

Variations or versions

Slight modifications

Adaptations

Original   Same work   New work

Cataloging Rules cut-off point

Same Expression   **New Expression**   **New Work**

IFLA Model: work, expression, manifestation, item

CONTENT work

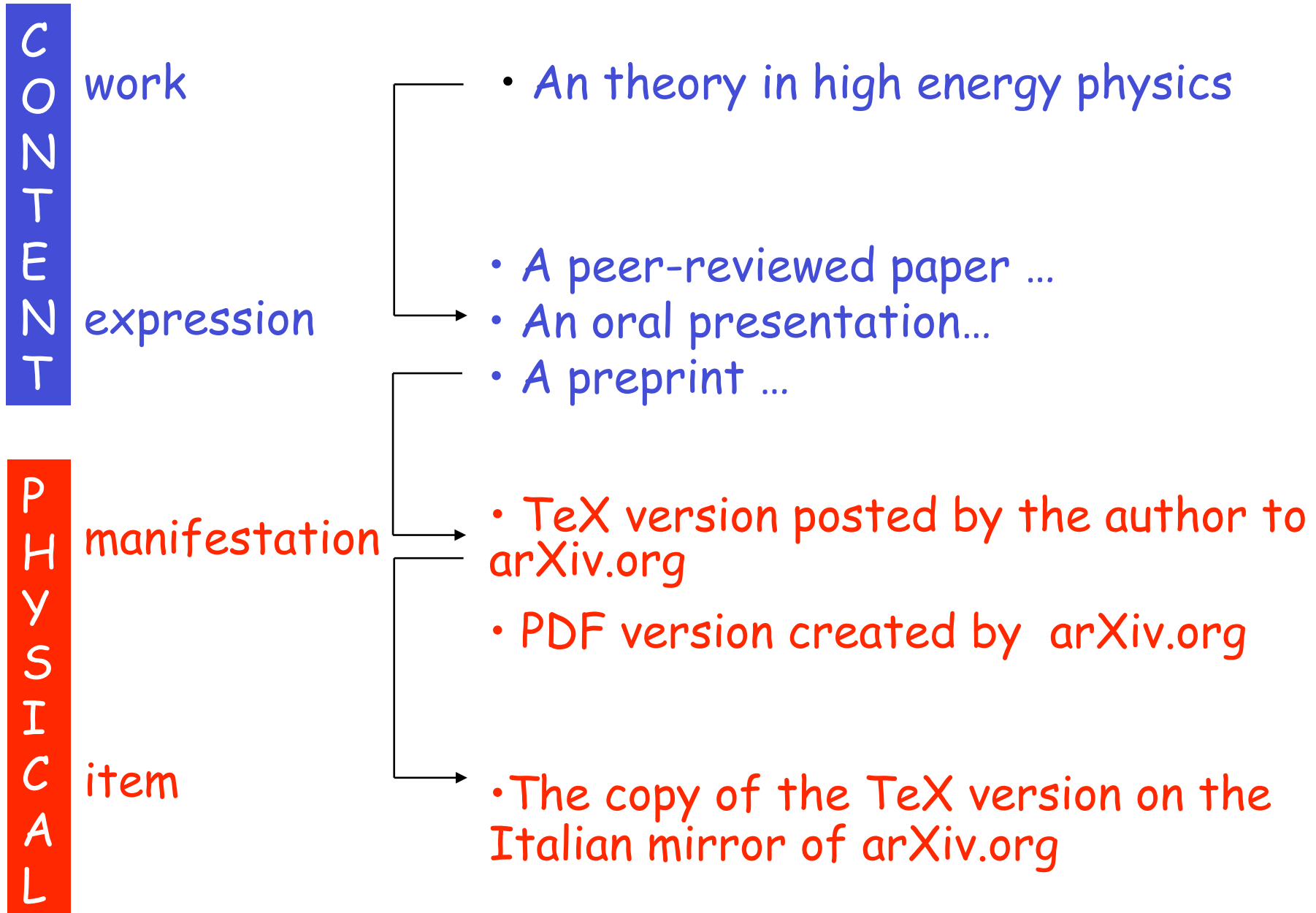- Psycho Killer

IFLA Model: work, expression, manifestation, item

C O N T E N T work   • Psycho Killer

                                • The score by David Byrne

expression   • The song as played by the Talking Heads

IFLA Model: work, expression, manifestation, item

CONTENT

work
- Psycho Killer

expression
- The score by David Byrne
- The song as played by the Talking Heads

PHYSICAL

manifestation
- Live "Stop Making Sense" performance in 1993 in Hollywood
- DVD of 1993 Performance
- The 45 RPM single released in 1979

IFLA Model: work, expression, manifestation, item

CONTENT

work
- Psycho Killer

expression
- The score by David Byrne
- The song as played by the Talking Heads

PHYSICAL

manifestation
- Live "Stop Making Sense" performance in 1993 in Hollywood
- DVD of 1993 Performance
- The 45 RPM single released in 1979

item
- Carl's personal DVD

IFLA Model: work, expression, manifestation, item

CONTENT

work — • An theory in high energy physics

expression —
- A peer-reviewed paper …
- An oral presentation…
- A preprint …

PHYSICAL

manifestation —
- TeX version posted by the author to arXiv.org
- PDF version created by arXiv.org

item —
- The copy of the TeX version on the Italian mirror of arXiv.org

Thinking about other relationships

Whole-Part Relationships

- **Components**
- **Aggregates**

Part-to-Part Relationships

- **Component to component**
  - **Sequential Relationship**
  - **Accompanying Relationship (or Companion Relationship)**
    - **Dependent**
    - **Independent**

Why should we care about all this?

- Matches a cognitive model of our information seeking and usage behavior
- Impacts intellectual property interests and laws
- Citations and reference linking
- Drives preservation decisions

Some real applications of FRBR

Browse | Search harry potter **GO** [Advanced]

Welcome

FictionFinder is a project of OCLC Research that provides access to millions of records for books, eBooks, and audio materials identified as fiction (including novels, short stories, dramas, and comics).

With FictionFinder, you can search these records by author, title, ISBN, and subject. You can also search book summaries and Search and Browse by genres, characters (real & fictional), settings (real & imaginary) and literary awards. And, when you discover an item you are interested in, FictionFinder can help you find a copy.

Subject Clouds | Dutch | English | French | Spanish

Actors Actresses Adultery Amnesia Angels Archaeologists Artists Assassins Boys Brothers Cats City and town life Clergy Composers Conspiracies Cookery Criminals Cults Death Depressions Dogs Dreams Ex-convicts Explorers Family Fathers and daughters Fathers and sons Foundlings Friendship Gay men Girls Governesses Grandmothers Heroes Hotelkeepers Immigrants Immortalism Inheritance and succession Islands Journalists Kidnapping Legislators Lesbians Mafia Magic Magicians Man-woman relationships Marriage Married women Millionaires Monsters Mormons Mothers and daughters Murder Murderers Nobility Novelists Nurses Opera Orphans Outlaws Physicians Pianists Pirates Poets Policewomen Presidents Prisoners Psychics Queens Reincarnation Revenge Sailors Scientists Sheriffs Sisters Slaves Soldiers Supernatural Terrorism Time travel Totalitarianism Twins Vampires Werewolves Widowers Widows Witches Wizards Women Young men Young women

* Faceted Application of Subject Terminology (FAST)

http://fictionfinder.oclc.org/index.html

Works display

Browse | **Search** [   ] **GO** [Advanced]

You searched: Basic Index for **harry** near **potter**

Previous 10 Works   Displaying 1 - 10 of 62   **Next 10 Works >>**   Jump to work: [1]
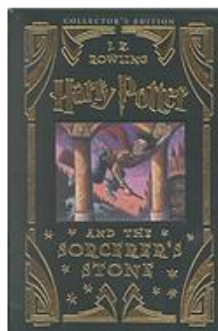
	Title / Author	Editions	Libraries ▼
1.	**Harry Potter and the Philosopher's Stone** / Rowling, J. K Rescued from the outrageous neglect of his aunt and uncle, a young boy with a great destiny proves his worth while attending Hogwarts School for Wizards and Witches.	162	12470
2.	**Harry Potter and the Chamber of Secrets** / Rowling, J. K When the Chamber of Secrets is opened again at the Hogwarts School for Witchcraft and Wizardry, second-year student Harry Potter finds himself in danger from a dark power that has once more been released on the school. In this highly celebrated sequel to: Harry Potter & the Sorcerer's Stone, Harry Potter spends a second, very eventful year at Hogwarts School for Witchcraft & Wizardry. During his second year at Hogwarts School of Witchcraft & Wizardry, Harry is baffled when he hears noises no one else can.	143	11903
3.	**Harry Potter and the Prisoner of Azkaban** / Rowling, J. K During his third year at Hogwarts School for Witchcraft and Wizardry, Harry Potter must confront the devious and dangerous wizard responsible for his parents' deaths.	117	10749
4.	**Harry Potter and the Goblet of Fire** / Rowling, J. K Fourteen-year-old Harry Potter joins the Weasleys at the Quidditch World Cup, then enters his fourth year at Hogwarts Academy where he is mysteriously entered in an unusual contest that challenges his wizarding skills, friendships and character, amid signs that an old enemy is growing stronger.	104	10126
5.	**Harry Potter and the Order of the Phoenix** / Rowling, J. K When the government of the magic world and authorities at Hogwarts School of Witchcraft and Wizardry refuse to believe in the growing threat of a freshly revived Lord Voldemort, fifteen-year-old Harry Potter finds support from his loyal friends in facing the evil wizard and other new terrors.	75	8904
6.	**Harry Potter and the Half-blood Prince** / Rowling, J. K Sixth-year Hogwarts student Harry Potter gains valuable insights into the boy Voldemort once was, even as his own world is transformed by maturing friendships, schoolwork assistance from an unexpected source, and devastating losses.	46	6846
7.	**A Muggle's Guide To the Wizarding World: Exploring the Harry Potter Universe** / Boyle, Fionna	1	299

Work/expression (editions) display