User Modelling and Archive Engineering

B. R. Barkstrom

Presentation at UNC School of Information and Library Science

October 10, 2014

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

1. Outline

This presentation wants to engage your thoughts by asking you to

- Carefully observe scientific data
- Then formulate answers to some questions

The Major Question Areas are

- Is the Sun's Energy Input to Earth's Climate Changing?
- What Kind of Users Interact With a Rock Core Archive?
- Can Users Find Data Without Keywords?
- How Do We Formulate a Model of User Activities?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

1. Outline

This presentation wants to engage your thoughts by asking you to

- Carefully observe scientific data
- Then formulate answers to some questions

The Major Question Areas are

- Is the Sun's Energy Input to Earth's Climate Changing?
- What Kind of Users Interact With a Rock Core Archive?
- Can Users Find Data Without Keywords?
- How Do We Formulate a Model of User Activities?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

1. Outline

This presentation wants to engage your thoughts by asking you to

- Carefully observe scientific data
- Then formulate answers to some questions

The Major Question Areas are

- Is the Sun's Energy Input to Earth's Climate Changing?
- What Kind of Users Interact With a Rock Core Archive?
- Can Users Find Data Without Keywords?
- How Do We Formulate a Model of User Activities?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

1. Outline

This presentation wants to engage your thoughts by asking you to

- Carefully observe scientific data
- Then formulate answers to some questions

The Major Question Areas are

- Is the Sun's Energy Input to Earth's Climate Changing?
- What Kind of Users Interact With a Rock Core Archive?
- Can Users Find Data Without Keywords?
- How Do We Formulate a Model of User Activities?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

1. Outline

This presentation wants to engage your thoughts by asking you to

- Carefully observe scientific data
- Then formulate answers to some questions

The Major Question Areas are

- Is the Sun's Energy Input to Earth's Climate Changing?
- What Kind of Users Interact With a Rock Core Archive?
- Can Users Find Data Without Keywords?
- How Do We Formulate a Model of User Activities?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

1. Outline

This presentation wants to engage your thoughts by asking you to

- Carefully observe scientific data
- Then formulate answers to some questions

The Major Question Areas are

- Is the Sun's Energy Input to Earth's Climate Changing?
- What Kind of Users Interact With a Rock Core Archive?
- Can Users Find Data Without Keywords?
- How Do We Formulate a Model of User Activities?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

1. Outline

This presentation wants to engage your thoughts by asking you to

- Carefully observe scientific data
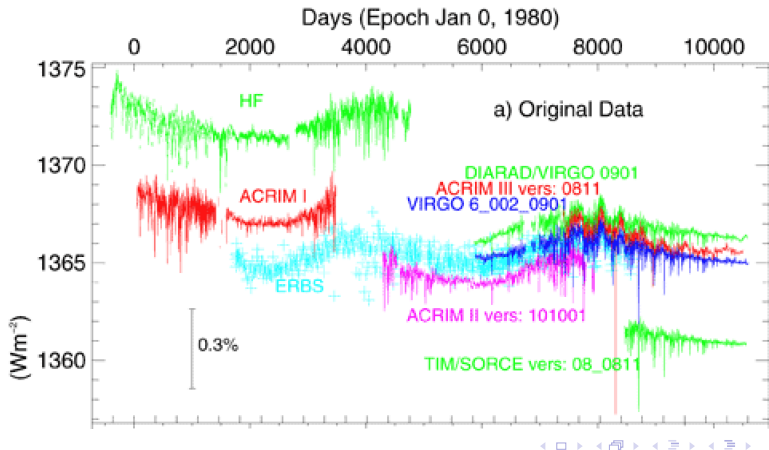- Then formulate answers to some questions

The Major Question Areas are

- Is the Sun's Energy Input to Earth's Climate Changing?
- What Kind of Users Interact With a Rock Core Archive?
- Can Users Find Data Without Keywords?
- How Do We Formulate a Model of User Activities?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

1. Is the Sun's Output Changing?

- Data From Several Satellite Instruments

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Question on the Sun's Output

- Jot down a note indicating whether you think these data show
  - Sun's output is increasing with time
  - Sun's output is more or less constant
  - Sun's output is decreasing with time
  - I need more information
- If you need more information, jot down what you think would help?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Question on the Sun's Output

- Jot down a note indicating whether you think these data show
  - Sun's output is increasing with time
  - Sun's output is more or less constant
  - Sun's output is decreasing with time
  - I need more information
- If you need more information, jot down what you think would help?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Some Contextual Ideas

Robertson, S. and Zaragoza, H., 2009: *The Probabilistic Relevance Framework: BM25 and Beyond* suggest data users have **Information Needs** (or **Search Intents**). We'll categorize these as

- **Desired Outcomes**: what would make the user feel his or her information needs were satisfied
- **Archive Output**: the objects the archive provides to try to help the user reach the Desired Outcome

- Outcomes may not be easy to quantify
- Output is the basis for engineering an archive's capacity to meet user's needs

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Some Contextual Ideas

Robertson, S. and Zaragoza, H., 2009: *The Probabilistic Relevance Framework: BM25 and Beyond* suggest data users have **Information Needs** (or **Search Intents**). We'll categorize these as

- **Desired Outcomes**: what would make the user feel his or her information needs were satisfied
- **Archive Output**: the objects the archive provides to try to help the user reach the Desired Outcome

- Outcomes may not be easy to quantify
- Output is the basis for engineering an archive's capacity to meet user's needs

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Some Contextual Ideas

Robertson, S. and Zaragoza, H., 2009: *The Probabilistic Relevance Framework: BM25 and Beyond* suggest data users have **Information Needs** (or **Search Intents**). We'll categorize these as

- **Desired Outcomes**: what would make the user feel his or her information needs were satisfied
- **Archive Output**: the objects the archive provides to try to help the user reach the Desired Outcome

- Outcomes may not be easy to quantify
- Output is the basis for engineering an archive's capacity to meet user's needs

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Some Contextual Ideas

Robertson, S. and Zaragoza, H., 2009: *The Probabilistic Relevance Framework: BM25 and Beyond* suggest data users have **Information Needs** (or **Search Intents**). We'll categorize these as

- **Desired Outcomes**: what would make the user feel his or her information needs were satisfied
- **Archive Output**: the objects the archive provides to try to help the user reach the Desired Outcome

- Outcomes may not be easy to quantify
- Output is the basis for engineering an archive's capacity to meet user's needs

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Some Contextual Ideas

Robertson, S. and Zaragoza, H., 2009: *The Probabilistic Relevance Framework: BM25 and Beyond* suggest data users have **Information Needs** (or **Search Intents**). We'll categorize these as

- **Desired Outcomes**: what would make the user feel his or her information needs were satisfied
- **Archive Output**: the objects the archive provides to try to help the user reach the Desired Outcome

- Outcomes may not be easy to quantify
- Output is the basis for engineering an archive's capacity to meet user's needs

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Applying These Ideas to the Solar Constant Data

- Initial impression of data
    - **Desired Outcome**: Reliable conclusion about whether the Sun's output was changing
    - **Archive Output**: Data from $\approx$10,000 days of observation (multiple files)
- Revised information need
    - **Desired Outcome**: Understanding the physical basis for correcting the data to get a better understood trend
    - **Archive Output**: Data **and** a mathematical model of each instrument, with derivations, algorithms, and source code
- How would you modify the archive's systems to *sense* the change in Desired Output and *respond appropriately* to the perceived change?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Applying These Ideas to the Solar Constant Data

- Initial impression of data
  - **Desired Outcome**: Reliable conclusion about whether the Sun's output was changing
  - **Archive Output**: Data from $\approx$10,000 days of observation (multiple files)
- Revised information need
  - **Desired Outcome**: Understanding the physical basis for correcting the data to get a better understood trend
  - **Archive Output**: Data **and** a mathematical model of each instrument, with derivations, algorithms, and source code
- How would you modify the archive's systems to *sense* the change in Desired Output and *respond appropriately* to the perceived change?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Applying These Ideas to the Solar Constant Data

- Initial impression of data
    - **Desired Outcome**: Reliable conclusion about whether the Sun's output was changing
    - **Archive Output**: Data from $\approx$10,000 days of observation (multiple files)
- Revised information need

    - **Desired Outcome:** Understanding the physical basis for correcting the data to get a better understood trend
    - **Archive Output:** Data **and** a mathematical model of each instrument, with derivations, algorithms, and source code

- How would you modify the archive's systems to *sense* the change in Desired Output and *respond appropriately* to the perceived change?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Applying These Ideas to the Solar Constant Data

- Initial impression of data
  - **Desired Outcome**: Reliable conclusion about whether the Sun's output was changing
  - **Archive Output**: Data from $\approx$10,000 days of observation (multiple files)
- Revised information need
  - **Desired Outcome**: Understanding the physical basis for correcting the data to get a better understood trend
  - **Archive Output**: Data **and** a mathematical model of each instrument, with derivations, algorithms, and source code
- How would you modify the archive's systems to *sense* the change in Desired Output and *respond appropriately* to the perceived change?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Applying These Ideas to the Solar Constant Data

- Initial impression of data
    - **Desired Outcome**: Reliable conclusion about whether the Sun's output was changing
    - **Archive Output**: Data from $\approx$10,000 days of observation (multiple files)
- Revised information need
    - **Desired Outcome**: Understanding the physical basis for correcting the data to get a better understood trend
    - **Archive Output**: Data **and** a mathematical model of each instrument, with derivations, algorithms, and source code
- How would you modify the archive's systems to *sense* the change in Desired Output and *respond appropriately* to the perceived change?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities
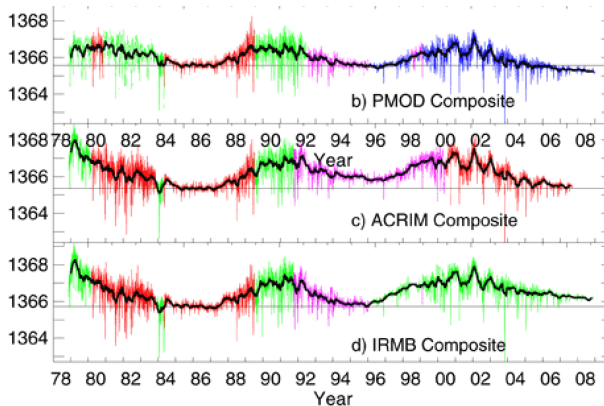
Applying These Ideas to the Solar Constant Data

- Initial impression of data
  - **Desired Outcome**: Reliable conclusion about whether the Sun's output was changing
  - **Archive Output**: Data from $\approx$10,000 days of observation (multiple files)
- Revised information need
  - **Desired Outcome**: Understanding the physical basis for correcting the data to get a better understood trend
  - **Archive Output**: Data **and** a mathematical model of each instrument, with derivations, algorithms, and source code
- How would you modify the archive's systems to *sense* the change in Desired Output and *respond appropriately* to the perceived change?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Applying These Ideas to the Solar Constant Data

- Initial impression of data
    - **Desired Outcome**: Reliable conclusion about whether the Sun's output was changing
    - **Archive Output**: Data from $\approx$10,000 days of observation (multiple files)
- Revised information need
    - **Desired Outcome**: Understanding the physical basis for correcting the data to get a better understood trend
    - **Archive Output**: Data **and** a mathematical model of each instrument, with derivations, algorithms, and source code
- How would you modify the archive's systems to *sense* the change in Desired Output and *respond appropriately* to the perceived change?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Which Corrections Produce the Best Estimate?

"Corrected Solar Constant" Data Estimated by Three Groups of Scientists

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

How Would You Cite the Sources of the "Corrected" Data?

Context-setting considerations:

- Journals only have room for a half page of references on a typical paper
- An investigator may want to devote particular attention to small subsets of data

Some suggestions:

- Make citation an archival *web service* that references sources for each data point
- Visualize selected data

Jot down your own ideas so we can discuss them.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

How Would You Cite the Sources of the "Corrected" Data?

Context-setting considerations:

- Journals only have room for a half page of references on a typical paper
- An investigator may want to devote particular attention to small subsets of data

Some suggestions:

- Make citation an archival *web service* that references sources for each data point
- Visualize selected data

Jot down your own ideas so we can discuss them.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

How Would You Cite the Sources of the "Corrected" Data?

Context-setting considerations:

- Journals only have room for a half page of references on a typical paper
- An investigator may want to devote particular attention to small subsets of data

Some suggestions:

- Make citation an archival *web service* that references sources for each data point
- Visualize selected data

Jot down your own ideas so we can discuss them.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

How Would You Cite the Sources of the "Corrected" Data?

Context-setting considerations:

- Journals only have room for a half page of references on a typical paper
- An investigator may want to devote particular attention to small subsets of data

Some suggestions:

- Make citation an archival *web service* that references sources for each data point
- Visualize selected data

Jot down your own ideas so we can discuss them.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

2. What Kinds of User Interact With a Rock Core Archive?

- Descriptions of users sometimes seem to treat all users as being "The Same"
- Users can differ in many ways
  - Age
  - Educational Attainment
  - Disciplinary Experience (years of on-the-job work)
  - Skill Sets (including math skills)

- Jot down any other categories of differences and examples of differences you can think of

B. R. Barkstrom   User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

2. What Kinds of User Interact With a Rock Core Archive?

- Descriptions of users sometimes seem to treat all users as being "The Same"
- Users can differ in many ways
  - Age
  - Educational Attainment
  - Disciplinary Experience (years of on-the-job work)
  - Skill Sets (including math skills)
- Jot down any other categories of differences and examples of differences you can think of

B. R. Barkstrom   User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

2. What Kinds of User Interact With a Rock Core Archive?

- Descriptions of users sometimes seem to treat all users as being "The Same"
- Users can differ in many ways
  - Age
  - Educational Attainment
  - Disciplinary Experience (years of on-the-job work)
  - Skill Sets (including math skills)
- Jot down any other categories of differences and examples of differences you can think of

B. R. Barkstrom   User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

2. What Kinds of User Interact With a Rock Core Archive?

- Descriptions of users sometimes seem to treat all users as being "The Same"
- Users can differ in many ways
    - Age
    - Educational Attainment
    - Disciplinary Experience (years of on-the-job work)
    - Skill Sets (including math skills)
- Jot down any other categories of differences and examples of differences you can think of

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

2. What Kinds of User Interact With a Rock Core Archive?

- Descriptions of users sometimes seem to treat all users as being "The Same"
- Users can differ in many ways
  - Age
  - Educational Attainment
  - Disciplinary Experience (years of on-the-job work)
  - Skill Sets (including math skills)
- Jot down any other categories of differences and examples of differences you can think of

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

2. What Kinds of User Interact With a Rock Core Archive?

- Descriptions of users sometimes seem to treat all users as being "The Same"
- Users can differ in many ways
    - Age
    - Educational Attainment
    - Disciplinary Experience (years of on-the-job work)
    - Skill Sets (including math skills)
- Jot down any other categories of differences and examples of differences you can think of

B. R. Barkstrom | User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

2. What Kinds of User Interact With a Rock Core Archive?

- Descriptions of users sometimes seem to treat all users as being "The Same"
- Users can differ in many ways
    - Age
    - Educational Attainment
    - Disciplinary Experience (years of on-the-job work)
    - Skill Sets (including math skills)
- Jot down any other categories of differences and examples of differences you can think of

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Rock Core Generation

Rock Cores Are Rocks Obtained by Drilling a Well

- Drilling crew uses a long hollow pipe that may be 1 to 2 km (about a mile) long.

- When the crew extracts the pipe from the well, they put the rocks in the pipe in boxes about 1 m (3 ft) long, making sure to keep the order of the layers.

- By law, part of the rock cores in the boxes must be deposited into a state-maintained rock core archive.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Rock Core Generation

Rock Cores Are Rocks Obtained by Drilling a Well

- Drilling crew uses a long hollow pipe that may be 1 to 2 km (about a mile) long.
- When the crew extracts the pipe from the well, they put the rocks in the pipe in boxes about 1 m (3 ft) long, making sure to keep the order of the layers.
- By law, part of the rock cores in the boxes must be deposited into a state-maintained rock core archive.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Rock Core Generation

Rock Cores Are Rocks Obtained by Drilling a Well

- Drilling crew uses a long hollow pipe that may be 1 to 2 km (about a mile) long.
- When the crew extracts the pipe from the well, they put the rocks in the pipe in boxes about 1 m (3 ft) long, making sure to keep the order of the layers.
- By law, part of the rock cores in the boxes must be deposited into a state-maintained rock core archive.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Store the Boxed Cores in the Archive's Stacks

Ms. Denise Hills, Program Director of Energy Investigations for the Alabama Geological Survey's Rock Core Archive Shows How the Archive Stores the Labelled Boxes in the Archive Stacks.

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities
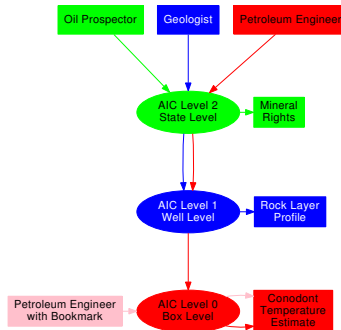
Experiences

Box Labels Record Metadata for the Cores They Contain

Some Rock Cores are Just Fragments That Need Their own Identification Within the Organization Provided by the Box Labels

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

A Rock Core Archive Has a Hierarchical Collection Structure

The Archive Information Collections (AIC's) Form a Hierarchy Used by Different User Communities in Different Ways

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Information Outcomes and Outputs

- Oil Prospector (informal "civilian" dress)
    - **Outcome**: townships with unsurveyed strata
    - **Output**: list of well permits including civil locations
- Geologist (academic dress – usually has all fingers and fingertips)

    - **Outcome**: 3D visualization of strata below surface
    - **Output**: Rock layer profiles in one or more wells

- Petroleum Engineer (managerial dress - white shirt and tie)

    - **Outcome**: untapped mineral resource-rich areas
    - **Output**: well strata with conodont samples

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Information Outcomes and Outputs

- Oil Prospector (informal "civilian" dress)
    - **Outcome**: townships with unsurveyed strata
    - **Output**: list of well permits including civil locations
- Geologist (academic dress – usually has all fingers and fingertips)

    - Outcome: 3D visualization of strata below surface
    - Output: Rock layer profiles in one or more wells

- Petroleum Engineer (managerial dress - white shirt and tie)

    - Outcome: untapped mineral resource-rich areas
    - Output: well strata with conodont samples

B. R. Barkstrom   User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Information Outcomes and Outputs

- Oil Prospector (informal "civilian" dress)
    - **Outcome**: townships with unsurveyed strata
    - **Output**: list of well permits including civil locations
- Geologist (academic dress – usually has all fingers and fingertips)
    - Outcome: 3D visualization of strata below surface
    - Output: Rock layer profiles in one or more wells
- Petroleum Engineer (managerial dress - white shirt and tie)

    - Outcome: untapped mineral resource-rich areas
    - Output: well strata with conodont samples

B. R. Barkstrom   User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Information Outcomes and Outputs

- Oil Prospector (informal "civilian" dress)
    - **Outcome**: townships with unsurveyed strata
    - **Output**: list of well permits including civil locations
- Geologist (academic dress – usually has all fingers and fingertips)
    - **Outcome**: 3D visualization of strata below surface
    - **Output**: Rock layer profiles in one or more wells
- Petroleum Engineer (managerial dress - white shirt and tie)

    - **Outcome**: untapped mineral resource-rich areas
    - **Output**: well strata with conodont samples

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Information Outcomes and Outputs

- Oil Prospector (informal "civilian" dress)
    - **Outcome**: townships with unsurveyed strata
    - **Output**: list of well permits including civil locations
- Geologist (academic dress – usually has all fingers and fingertips)
    - **Outcome**: 3D visualization of strata below surface
    - **Output**: Rock layer profiles in one or more wells
- Petroleum Engineer (managerial dress - white shirt and tie)

    - **Outcome**: untapped mineral resource-rich areas
    - **Output**: well strata with conodont samples

B. R. Barkstrom   User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Information Outcomes and Outputs

- Oil Prospector (informal "civilian" dress)
    - **Outcome**: townships with unsurveyed strata
    - **Output**: list of well permits including civil locations
- Geologist (academic dress – usually has all fingers and fingertips)
    - **Outcome**: 3D visualization of strata below surface
    - **Output**: Rock layer profiles in one or more wells
- Petroleum Engineer (managerial dress - white shirt and tie)

    - **Outcome**: untapped mineral resource-rich areas
    - **Output**: well strata with conodont samples

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Vocabulary (Dialect) and Customs

- Oil Prospector
  - **Vocabulary**: Civil boundaries (townships) and English Units
  - **Customs**: Seeks to be indistinguishable from other people to avoid having other mineral rights prospectors recognize interest in unsurveyed areas
- Geologist
  - **Vocabulary**: Stratigraphic and Geological Features; Metric Units
  - **Customs**: Writes papers and judges professional status on this basis
- Petroleum Engineer
  - **Vocabulary**: Professional Geological Vocabulary With Managerial and Financial Overlay
  - **Customs**: Reads and writes company records

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Vocabulary (Dialect) and Customs

- Oil Prospector
  - **Vocabulary**: Civil boundaries (townships) and English Units
  - **Customs**: Seeks to be indistinguishable from other people to avoid having other mineral rights prospectors recognize interest in unsurveyed areas
- Geologist
  - **Vocabulary**: Stratigraphic and Geological Features; Metric Units
  - **Customs**: Writes papers and judges professional status on this basis
- Petroleum Engineer
  - **Vocabulary**: Professional Geological Vocabulary With Managerial and Financial Overlay
  - **Customs**: Reads and writes company records

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Vocabulary (Dialect) and Customs

- Oil Prospector
    - **Vocabulary**: Civil boundaries (townships) and English Units
    - **Customs**: Seeks to be indistinguishable from other people to avoid having other mineral rights prospectors recognize interest in unsurveyed areas
- Geologist
    - **Vocabulary**: Stratigraphic and Geological Features; Metric Units
    - **Customs**: Writes papers and judges professional status on this basis
- Petroleum Engineer
    - **Vocabulary**: Professional Geological Vocabulary With Managerial and Financial Overlay
    - **Customs**: Reads and writes company records

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Vocabulary (Dialect) and Customs

- Oil Prospector
  - **Vocabulary**: Civil boundaries (townships) and English Units
  - **Customs**: Seeks to be indistinguishable from other people to avoid having other mineral rights prospectors recognize interest in unsurveyed areas
- Geologist
  - **Vocabulary**: Stratigraphic and Geological Features; Metric Units
  - **Customs**: Writes papers and judges professional status on this basis
- Petroleum Engineer
  - **Vocabulary**: Professional Geological Vocabulary With Managerial and Financial Overlay
  - **Customs**: Reads and writes company records

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Vocabulary (Dialect) and Customs

- Oil Prospector
    - **Vocabulary**: Civil boundaries (townships) and English Units
    - **Customs**: Seeks to be indistinguishable from other people to avoid having other mineral rights prospectors recognize interest in unsurveyed areas
- Geologist
    - **Vocabulary**: Stratigraphic and Geological Features; Metric Units
    - **Customs**: Writes papers and judges professional status on this basis
- Petroleum Engineer
    - **Vocabulary**: Professional Geological Vocabulary With Managerial and Financial Overlay
    - **Customs**: Reads and writes company reports

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Each User Community Has Its Own Vocabulary (Dialect) and Customs

- Oil Prospector
  - **Vocabulary**: Civil boundaries (townships) and English Units
  - **Customs**: Seeks to be indistinguishable from other people to avoid having other mineral rights prospectors recognize interest in unsurveyed areas
- Geologist
  - **Vocabulary**: Stratigraphic and Geological Features; Metric Units
  - **Customs**: Writes papers and judges professional status on this basis
- Petroleum Engineer
  - **Vocabulary**: Professional Geological Vocabulary With Managerial and Financial Overlay
  - **Customs**: Reads and writes company reports

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Questions: How Does an Archive Accommodate User Community Diversity

- Should an archive have one path to its holdings or should it have tailored paths that adapt to the recognizable communities it has to deal with?

- If it chooses to tailor paths, how does it recognize to which community a particular user belongs?

- Jot down your suggestions on these questions.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Questions: How Does an Archive Accommodate User Community Diversity

- Should an archive have one path to its holdings or should it have tailored paths that adapt to the recognizable communities it has to deal with?

- If it chooses to tailor paths, how does it recognize to which community a particular user belongs?

- Jot down your suggestions on these questions.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Questions: How Does an Archive Accommodate User Community Diversity

- Should an archive have one path to its holdings or should it have tailored paths that adapt to the recognizable communities it has to deal with?

- If it chooses to tailor paths, how does it recognize to which community a particular user belongs?

- Jot down your suggestions on these questions.

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

3. Can Users Find Data Without Keywords

- Many of the approaches for finding data come from experience with document retrieval.
- Documents are usually text based.
- It's natural to think of using text based tools
    - Keyword Phrases
    - Controlled Vocabularies
    - Statistical Ranking Based on Keywords From Many Users
- Jot down other tools that use texual searches
- What kinds of problems can you see arising from **Semantic Heterogeneity**?

B. R. Barkstrom   User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

3. Can Users Find Data Without Keywords

- Many of the approaches for finding data come from experience with document retrieval.

- Documents are usually text based.

- It's natural to think of using text based tools

    - Keyword Phrases

    - Controlled Vocabularies

    - Statistical Ranking Based on Keywords From Many Users

- Jot down other tools that use texual searches

- What kinds of problems can you see arising from **Semantic Heterogeneity**?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

3. Can Users Find Data Without Keywords

- Many of the approaches for finding data come from experience with document retrieval.
- Documents are usually text based.
- It's natural to think of using text based tools
  - Keyword Phrases
  - Controlled Vocabularies
  - Statistical Ranking Based on Keywords From Many Users
- Jot down other tools that use texual searches
- What kinds of problems can you see arising from **Semantic Heterogeneity**?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

3. Can Users Find Data Without Keywords

- Many of the approaches for finding data come from experience with document retrieval.
- Documents are usually text based.
- It's natural to think of using text based tools
  - Keyword Phrases
  - Controlled Vocabularies
  - Statistical Ranking Based on Keywords From Many Users
- Jot down other tools that use texual searches
- What kinds of problems can you see arising from **Semantic Heterogeneity**?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

3. Can Users Find Data Without Keywords

- Many of the approaches for finding data come from experience with document retrieval.
- Documents are usually text based.
- It's natural to think of using text based tools
  - Keyword Phrases
  - Controlled Vocabularies
  - Statistical Ranking Based on Keywords From Many Users
- Jot down other tools that use texual searches
- What kinds of problems can you see arising from **Semantic Heterogeneity**?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

3. Can Users Find Data Without Keywords

- Many of the approaches for finding data come from experience with document retrieval.
- Documents are usually text based.
- It's natural to think of using text based tools
  - Keyword Phrases
  - Controlled Vocabularies
  - Statistical Ranking Based on Keywords From Many Users
- Jot down other tools that use texual searches
- What kinds of problems can you see arising from **Semantic Heterogeneity**?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

3. Can Users Find Data Without Keywords

- Many of the approaches for finding data come from experience with document retrieval.
- Documents are usually text based.
- It's natural to think of using text based tools
  - Keyword Phrases
  - Controlled Vocabularies
  - Statistical Ranking Based on Keywords From Many Users
- Jot down other tools that use texual searches
- What kinds of problems can you see arising from **Semantic Heterogeneity**?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

3. Can Users Find Data Without Keywords

- Many of the approaches for finding data come from experience with document retrieval.
- Documents are usually text based.
- It's natural to think of using text based tools
  - Keyword Phrases
  - Controlled Vocabularies
  - Statistical Ranking Based on Keywords From Many Users
- Jot down other tools that use texual searches
- What kinds of problems can you see arising from **Semantic Heterogeneity**?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Interesting Example

- NOAA has a collection of aerial photographs for helping with damage assessment after major storms (Katrina, Sandy); see http://storms.ngs.noaa.gov/eri.page/index.html
- Within 2 or 3 days after the storm (when there are few clouds), NOAA sends an aircraft with a camera to obtain images of areas with severe damage
- There are many thousands of images
- Users can obtain individual full-resolution jpg images, individual thumbnails, or files tha zip up many images along portions of a flight path
- Jot down your idea of a strategy for helping a user find an appropriate search output
- Can you think of a way of helping the user that doesn't use keywords or rankings or latitude and longitude?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Interesting Example

- NOAA has a collection of aerial photographs for helping with damage assessment after major storms (Katrina, Sandy); see http://storms.ngs.noaa.gov/eri.page/index.html
- Within 2 or 3 days after the storm (when there are few clouds), NOAA sends an aircraft with a camera to obtain images of areas with severe damage
- There are many thousands of images
- Users can obtain individual full-resolution jpg images, individual thumbnails, or files tha zip up many images along portions of a flight path
- Jot down your idea of a strategy for helping a user find an appropriate search output
- Can you think of a way of helping the user that doesn't use keywords or rankings or latitude and longitude?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Interesting Example

- NOAA has a collection of aerial photographs for helping with damage assessment after major storms (Katrina, Sandy); see http://storms.ngs.noaa.gov/eri.page/index.html
- Within 2 or 3 days after the storm (when there are few clouds), NOAA sends an aircraft with a camera to obtain images of areas with severe damage
- There are many thousands of images
- Users can obtain individual full-resolution jpg images, individual thumbnails, or files tha zip up many images along portions of a flight path
- Jot down your idea of a strategy for helping a user find an appropriate search output
- Can you think of a way of helping the user that doesn't use keywords or rankings or latitude and longitude?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Interesting Example

- NOAA has a collection of aerial photographs for helping with damage assessment after major storms (Katrina, Sandy); see http://storms.ngs.noaa.gov/eri.page/index.html
- Within 2 or 3 days after the storm (when there are few clouds), NOAA sends an aircraft with a camera to obtain images of areas with severe damage
- There are many thousands of images
- Users can obtain individual full-resolution jpg images, individual thumbnails, or files tha zip up many images along portions of a flight path
- Jot down your idea of a strategy for helping a user find an appropriate search output
- Can you think of a way of helping the user that doesn't use keywords or rankings or latitude and longitude?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Interesting Example

- NOAA has a collection of aerial photographs for helping with damage assessment after major storms (Katrina, Sandy); see http://storms.ngs.noaa.gov/eri.page/index.html
- Within 2 or 3 days after the storm (when there are few clouds), NOAA sends an aircraft with a camera to obtain images of areas with severe damage
- There are many thousands of images
- Users can obtain individual full-resolution jpg images, individual thumbnails, or files tha zip up many images along portions of a flight path
- Jot down your idea of a strategy for helping a user find an appropriate search output
- Can you think of a way of helping the user that doesn't use keywords or rankings or latitude and longitude?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Interesting Example

- NOAA has a collection of aerial photographs for helping with damage assessment after major storms (Katrina, Sandy); see http://storms.ngs.noaa.gov/eri.page/index.html
- Within 2 or 3 days after the storm (when there are few clouds), NOAA sends an aircraft with a camera to obtain images of areas with severe damage
- There are many thousands of images
- Users can obtain individual full-resolution jpg images, individual thumbnails, or files tha zip up many images along portions of a flight path
- Jot down your idea of a strategy for helping a user find an appropriate search output
- Can you think of a way of helping the user that doesn't use keywords or rankings or latitude and longitude?

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Some Observations on the NOAA ERI Example

- Usual approach involves a search engine whose designers assume they can guess what the user needs without asking the user directly

- A librarian suggested that a real reference librarian would have a **reference interview** to provide a search context and suggest resources to match the search to the perceived user needs

B. R. Barkstrom   User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Some Observations on the NOAA ERI Example

- Usual approach involves a search engine whose designers assume they can guess what the user needs without asking the user directly
- A librarian suggested that a real reference librarian would have a **reference interview** to provide a search context and suggest resources to match the search to the perceived user needs

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Alternative Model

- A standard Artifical Intelligence text [Russell and Norvig, 2010] suggests treating the problem as a **two-agent cooperative game**
  - The **Search Engine** doesn't know what the user wants – but *knows the archive's contents*
  - The **User** may not know what the archive contains – but *will know when his information need is satisfied* (or not)
  - Players take turns (Search Engine suggests possible paths; User chooses one)

- Jot down your approach to designing a Web Interface that could use this approach
  - *Hint*: treat each Web page as a node in a graph;
  - Each page provides hints with URL's that link to other pages

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Alternative Model

- A standard Artifical Intelligence text [Russell and Norvig, 2010] suggests treating the problem as a **two-agent cooperative game**
    - The **Search Engine** doesn't know what the user wants – but *knows the archive's contents*
    - The **User** may not know what the archive contains – but *will know when his information need is satisfied* (or not)
    - Players take turns (Search Engine suggests possible paths; User chooses one)

- Jot down your approach to designing a Web Interface that could use this approach

    - *Hint*: treat each Web page as a node in a graph;
    - Each page provides hints with URL's that link to other pages

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Alternative Model

- A standard Artifical Intelligence text [Russell and Norvig, 2010] suggests treating the problem as a **two-agent cooperative game**
    - The **Search Engine** doesn't know what the user wants – but *knows the archive's contents*
    - The **User** may not know what the archive contains – but *will know when his information need is satisfied* (or not)
    - Players take turns (Search Engine suggests possible paths; User chooses one)

- Jot down your approach to designing a Web Interface that could use this approach

    - *Hint*: treat each Web page as a node in a graph;
    - Each page provides hints with URL's that link to other pages

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Alternative Model

- A standard Artifical Intelligence text [Russell and Norvig, 2010] suggests treating the problem as a **two-agent cooperative game**
  - The **Search Engine** doesn't know what the user wants – but *knows the archive's contents*
  - The **User** may not know what the archive contains – but *will know when his information need is satisfied* (or not)
  - Players take turns (Search Engine suggests possible paths; User chooses one)

- Jot down your approach to designing a Web Interface that could use this approach

    - *Hint*: treat each Web page as a node in a graph;
    - Each page provides hints with URL's that link to other pages

B. R. Barkstrom   User Modeling

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Alternative Model

- A standard Artifical Intelligence text [Russell and Norvig, 2010] suggests treating the problem as a **two-agent cooperative game**
    - The **Search Engine** doesn't know what the user wants – but *knows the archive's contents*
    - The **User** may not know what the archive contains – but *will know when his information need is satisfied* (or not)
    - Players take turns (Search Engine suggests possible paths; User chooses one)

- Jot down your approach to designing a Web Interface that could use this approach
    - *Hint*: treat each Web page as a node in a graph;
    - Each page provides hints with URL's that link to other pages

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Alternative Model

- A standard Artifical Intelligence text [Russell and Norvig, 2010] suggests treating the problem as a **two-agent cooperative game**
    - The **Search Engine** doesn't know what the user wants – but *knows the archive's contents*
    - The **User** may not know what the archive contains – but *will know when his information need is satisfied* (or not)
    - Players take turns (Search Engine suggests possible paths; User chooses one)

- Jot down your approach to designing a Web Interface that could use this approach
    - *Hint*: treat each Web page as a node in a graph;
    - Each page provides hints with URL's that link to other pages

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

An Alternative Model

- A standard Artifical Intelligence text [Russell and Norvig, 2010] suggests treating the problem as a **two-agent cooperative game**
    - The **Search Engine** doesn't know what the user wants – but *knows the archive's contents*
    - The **User** may not know what the archive contains – but *will know when his information need is satisfied* (or not)
    - Players take turns (Search Engine suggests possible paths; User chooses one)

- Jot down your approach to designing a Web Interface that could use this approach
    - *Hint*: treat each Web page as a node in a graph;
    - Each page provides hints with URL's that link to other pages

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

4. How Do We Formulate a Model of User Activities?

The basic components:

- **User State Description** showing the states a user goes through in searching
- Archive Inventory showing which parts of the archive's collect each activity interacts with
- **User Community Demographics and Characteristics** tied to the properties of the objects in the archive's inventory

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

4. How Do We Formulate a Model of User Activities?

The basic components:

- **User State Description** showing the states a user goes through in searching
- **Archive Inventory** showing which parts of the archive's collect each activity interacts with
- User Community Demographics and Characteristics tied to the properties of the objects in the archive's inventory

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

4. How Do We Formulate a Model of User Activities?

The basic components:

- **User State Description** showing the states a user goes through in searching

- **Archive Inventory** showing which parts of the archive's collect each activity interacts with

- **User Community Demographics and Characteristics** tied to the properties of the objects in the archive's inventory

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Insights From User Modelling

- These components form a **finite state model** for user interactions
  - A steady-state model of this type can be a **Markov model**
  - A dynamic model provides a **Monte Carlo model**
  - Either form can quantify the quantitative resource load for archive output
  - Either form can help sharpen up perceptions of how to satisfy user information needs and identify bottlenecks to user satisfaction

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Insights From User Modelling

- These components form a **finite state model** for user interactions
    - A steady-state model of this type can be a **Markov model**
    - A dynamic model provides a **Monte Carlo model**
    - Either form can quantify the quantitative resource load for archive output
    - Either form can help sharpen up perceptions of how to satisfy user information needs and identify bottlenecks to user satisfaction

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Insights From User Modelling

- These components form a **finite state model** for user interactions
    - A steady-state model of this type can be a **Markov model**
    - A dynamic model provides a **Monte Carlo model**
    - Either form can quantify the quantitative resource load for archive output
    - Either form can help sharpen up perceptions of how to satisfy user information needs and identify bottlenecks to user satisfaction

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Insights From User Modelling

- These components form a **finite state model** for user interactions
    - A steady-state model of this type can be a **Markov model**
    - A dynamic model provides a **Monte Carlo model**
    - Either form can quantify the quantitative resource load for archive output
    - Either form can help sharpen up perceptions of how to satisfy user information needs and identify bottlenecks to user satisfaction

Experiences

Outline
Solar Constant
Rock Cores
Non-Keyword Searches
Formulating a Model of User Activities

Insights From User Modelling

- These components form a **finite state model** for user interactions
    - A steady-state model of this type can be a **Markov model**
    - A dynamic model provides a **Monte Carlo model**
    - Either form can quantify the quantitative resource load for archive output
    - Either form can help sharpen up perceptions of how to satisfy user information needs and identify bottlenecks to user satisfaction