The Problem

Problem

- paramount to the success of your effort
- Must be stated precisely
- must address an important question
- must advance existing knowledge
- must be grounded in objective reality
- must hint at the possible solutions

How to formulate an important and useful research problem?

- First need to be knowledgeable about your topic of interest
  - read the literature - most likely it will tell you what needs to be done
  - especially workshop and conference papers in the area of research
- attend professional conferences
- seek the advice of experts

Characteristics of Research Problems

- Research project not for self-enlightenment
- Simply comparing 2 data sets not sufficient
- Simply computing correlations between data sets not sufficient
  - Need to ask why correlation exists
- Problems resulting in “yes” or “no” answers not good research problems
  - Need to focus on the “what” and the “why”

Possible Avenues for Research Problems

- Address the suggestions for future research that other researchers have offered
- replicate a project in a different setting or with a different population
- apply an existing perspective to a new situation
- challenge other research findings

Stating Your Research Problem

- After you have figured out what you are going to focus your effort on, you must state the problem clearly and completely
- examples of poorly stated problems:
  - Security on wireless products
  - Metrics for large systems
- lack clarity, no context, no reachable goals

In every research endeavor

- state the context/motivation for the problem
- state the research questions
- formulate hypothesis aimed at solving the problem
- delimit the research
- define the terms and concepts
- state the assumptions

Hypotheses

- Tentative propositions set forth to assist in guiding the investigation of the problem or to provide possible explanation for the observations made

Problem Statement

Find Your Problem

- By April 18, submit a clear, precise statement of a computer science problem for research.
- Guidelines:
  - problem is stated in complete, grammatical sentences
  - is clear how the area of study will be limited or focused
  - is more than a simple exercise in gathering information, answering a yes/no question or making a simple comparison
  - includes a discussion of methods and approaches to verify the hypotheses
- Complete the worksheet on pp. 60-61 of Practical Research.

Purpose of the Problem Statement

- Represents the reasons/motivation behind your proposal (based on the specific domain of study).
- It specifies the conditions you want to change or the gaps in existing knowledge you intend to fill (this is the specification of the research problem).
- Should be supported by evidence.
- Specifies your hypothesis that suggests a solution to the problem.
- Shows your familiarity with prior research on the topic and why it needs to be extended.
- Even if the problem is obvious, your reviewers want to know how clearly you can state it.

Key Questions to Answer in Your Problem Statement

1. Demonstrate a precise understanding of and the motivation behind the problem you are attempting to solve?
2. Clearly convey the focus of your project early in the narrative?
3. Indicate the relationship of your project to a larger set of problems and justify why your particular focus has been chosen?
4. Demonstrate that your hypothesis is supported by evidence and observations
5. Demonstrate that your problem is feasible to solve and that your experimental design is appropriate for validating your hypothesis?
6. Make others what to read it further?

Writing Tips for Problem Statement

- Do not paint the problem in general terms:
  - "little is known about .."
  - "no research has dealt with .."
- Usually arguing for something that isn't makes for a weak need statement
- Instead explain the consequences of the information void

Refine Your Problem Statement

- 1. Complete the checklist on page 50 of Practical Research.
- 2. Think about sub-problems and further delineate your statement.
- 3. Start completing the checklist on pp. 60-61, then go back to your problem statement/abstract and revise as necessary.

Examples of Problem Statements

MoJo: A Distance Metric for Software Clustering

The software clustering problem has attracted much attention recently, since it is an integral part of the process of reverse engineering large software systems. A key problem in this research is the difficulty in comparing different approaches in an objective fashion. **[Needs to say in more detail what the difficulty is]**
We propose a metric that calculates a distance between two partitions of the same set of software resources. We hypothesize that this metric can be used to effectively evaluate the similarity of two different decompositions of a software system.
We begin by introducing our model and present a heuristic algorithm that calculates the distance in an efficient fashion. We evaluate the performance of the algorithm and the effectiveness of the metric….
[Need to say more about the experiments and how they might be used to validate the hypothesis]

Task-Oriented Pattern Discovery for Predictive Web User Modeling

An essential task in building personalized and adaptive systems is the automatic discovery of predictive models for user behavior. Existing approaches, such as clustering, correlation analysis, and association discovery, tend to generate shallow patterns which do not capture the full complexity of users' online behavior. Nor can the generated patterns explain the users' underlying interests which lead to specific types of behavioral patterns. To better capture users' underlying interests or information needs, we introduce the notion of "task". A task is a set or sequence of actions which are likely to be performed commonly by users in order to meet a specific information need or perform a specific function. These tasks are not directly visible, but can be captured and characterized either by a combination of users' interactions with the site and the site's content and structure. (cont.)

Task-Oriented Pattern Discovery for Predictive Web User Modeling (cont.)

We hypothesize that patterns discovered at the task level can provide a better understanding of users' underlying interests, which in turn, can lead to better predictive models. We propose a framework for “Task-Oriented Web User Modeling”. We intend to use probabilistic latent variable modeling to automatically discover and quantify user “tasks” and task-level patterns from users’ navigation data, as well as from Web site's content and structure data. Based on this framework, we will propose a novel personalization approach, based on the maximum entropy principle, which allows for a seamless integration of content-based and usage-based task-level patterns. We will perform experiments on real Web usage data and movie rating data to validate that the proposed approach results in more accurate and flexible predictive models. **[Need more on metrics and experimental design]**

Personalization in Folksonomies Based on Tag Clustering

Collaborative tagging systems, sometimes referred to as "folksonomies," enable Internet users to annotate or search for resources using custom labels (tags) instead of being restricted by pre-defined navigational or conceptual hierarchies. However, the flexibility of tagging brings with it certain costs. Because users are free to apply any tag to any resource, tagging systems contain large numbers of redundant, ambiguous, and idiosyncratic tags which can render resource discovery difficult.

We believe that data mining techniques such as clustering can be used to ameliorate this problem by reducing noise in the data and identifying trends. In particular, discovered tag clusters based on their common occurrences across resources can be used to tailor and personalize the system's output to a user based on the user's tagging behavior. (cont….)

Personalization in Folksonomies Based on Tag Clustering (cont.)

A personalized view can overcome ambiguity and idiosyncratic tag assignment, presenting users with tags and resources that correspond more closely to their intent.

Specifically, we will examine unsupervised clustering methods for extracting commonalities between tags, and use the discovered clusters as intermediaries between a user's profile and resources in order to tailor the results of search to the user's interests.

We validate this approach through extensive evaluation of proposed personalization algorithm and the underlying clustering techniques using data from two real collaborative tagging Web sites. **[Need to say more about the expriments and why they are appropriate]**

Cheat Sheet / Algorithm for papers/abstract/proposals

All should have the following elements in this order:

1. The general case / problem
2. What others have done
3. What's missing / where is the gap
4. Our solution (or hypothesis, if it is a proposal) and why it fills the gap
5. Specific results (or research design, if it is a proposal)

“The intensity of the conviction that a hypothesis is true has no bearing on whether it is true or not.”

P.B. Medawar

Advice to a Young Scientist

“The great tragedy of science, the slaying of a beautiful hypothesis by an ugly fact.”

T.H. Huxley

Biogenesis and Abiogenesis

“Mankind only sets itself such problems as it can solve, since closer examination will always reveal that the problem itself only arises when the material conditions for its solution are already present or in the process of formation.”

-- Karl Marx, 1859