# Computer Knowledge Taxonomy

## By organization:

### Single page

### Slideshow

## By authors:

### Official

### Expert

### Novice

## By media types:

### Text-only

### Text and Images

### Screencast

## By medium:

### Book

### Blog

### Forum

### Mailing list

## By areas:

### Command line

### Programming

### GUI

# Knowledge Life Cycle

How each stage of the knowledge life cycle has some need?

Creation, aggregating, organizing, querying, retrieving, visualizing, consuming

(List the needs in the order of importance)

Two major needs. First major need for research on a specialized system. More technical issues. Second major need to cover a wide range of demographics. More human issues.

# Two Major Needs:

## A specialized system to access computer know-how knowledge. There are needs for research in each stage of the knowledge life cycle. Understand the different kinds, forms, types how this kind of knowledge is on the web

## Cover a wide range of user groups

# Need for a specialized dataset (aggregating)

## Problem:

No specialized knowledge base and search service for this important task

Use google, but, first finds site, then search within the site

Use book, but

Use one company’s site, but

## Goal:

Develop a crawling scheme that focus on useful computer tutorial materials.

Show the density of useful information is much higher than a strong man alternative.

## Method:

Shortcut: Use image search engine, take words from a corpus such as a book. Mix in words related to software name. Keep those with screenshots. Train a visual detector for screenshot images.

Brute-force way: Systematically crawl major tutorial websites

Filter useless pages that are not related to computer knowledge. Train a classifier for detecting pages that are computer knowledge.

# Need for multi-modal way to index (offline indexing)

## Problem:

visual algorithms are built for specific domains. it is unclear whether that can be applied directly to our problem.

vision algorithm has not been designed to consider visual features of software program

## Goal:

Specialized, multi-modal indexing scheme. Compared to keyword only and text only. Develop a scheme that uses images to create context and evaluate against a sizable dataset and show statistical significance more relevant results than start-of-the-art methods, when evaluated by human users. First page result. Above the fold result. Recall and precision.

## Method:

Independent of query. Use local features plus geometry. Develop GUI specific features. Index tutorials based on words and images. Images also based on OCR and based on visual features. Classify tutorials based on types, taxonomy, OS types, walkthrough, forum, blog. Authority. Trust. Maybe base on comments, and do simple sentiment analysis.

# Need for more contextually and semantically relevant results (online retrieving)

## Problem:

simply applying cbir won't work because text contents may not be relevant

simply applying ir won't work because image may not be relevant

## Goal:

after some candidates are retrieved, how to rank them?

## Method:

query dependent. rank results by both visual and textual relevance. Identify many features. Learn ranking, using RankSVM. Improve visual search. Provide faceted search function.

# Need for an easier and more intuitive input method to specify queries (querying)

## Problem:

Hard to come up with the right keywords.

Too many things need to be specified. OS. Application Name, window title, what the users want to know.

Keywords can be ambiguous

## Goal:

Develop a method that are statistically significant subjectively easier, takes less time compared to current methods using keywords or browsing outlines. More success rate of finding the desired item.

## Method:

allow users to capture sceenshot as query. Allow users to type keywords as query. Use Java to provide cross-platform applicability. Allow users to take multiple screenshots as query.

# Need for a quicker and more informed way to judge the relevancy from preview (visualizing)

## Problem:

Text only excerpt is not enough

Image search presentation is not informative enough. Only tell you the image is correct, nothing else.

## Goal:

New scheme to present excerpts

## Method:

show image excerpt. Show screenshot in context to let user know the context, what words are before and after that.

# Need for a more effective, convenient, interactive way to follow tutorials (consuming)

## Problem:

Can not random access

Need to switch back and forth.

Hard to synchronize the current program and the browser

## Goal:

Develop a synchronized tutorial viewing tool.

## Method:

monitor the entire screen, matching the screen to the images in a tutorial article. Automatically scroll the page to that image. Allow users to search for content by image. Browser function called Find by image.

# Needs for a cheaper and more accessible way to create interactive tutorials (creating, expanding the corpus)

## Problem:

Can not access API

Hard to anticipate problems before software is shipped.

## Goal:

Third-party content provider to develop tutorials. Develop a framework requires no API and can work on all software and on all platforms. Develop an automatic generator that produces results that can be cheaper to correct by humans and achieve comparable results, than generating interactive tutorials from scratch. Cheaper measured by man hours.

## Method:

Allow content providers to attach arbitrary information to programs by images.

Establishing visual links.

Submit an url to the system. And it downloads all the image and index the page.

# Major Need: Needs for better support for different demographic groups

## Problem

### Novice:

### Expert:

### Senior:

### Children:

## Goal:

Develop an end-to-end support system tested with seniors and children. Show potential of customization to for special groups. All the same goals will be applied to seniors too. Participatory design. Evaluated with seniors. Orthogonal to the other needs. We will identify at least two major tasks and study whether seniors can do better using the proposed system. We develop the system for general public. The benefits to seniors and children are even greater.