


# Camcrawler

*Webcam Search Engine*

Ricky Bargreen, Flavio Pfaffhauser,  
Forrest Vines, Yih Sun Khoo



# Project Overview

## Initial Idea:

- Crawl the web, download pages containing webcams
- Classify pages
- Extract data from webcams
  - location
  - video type
  - description
- Web-based search interface

## Final Project:

- Same as initial idea...

# Project Components

## Crawler

- Nutch
- Start from seeds
- Download tons of pages

## Classifier

- Keyword Filter (good and bad)
- Rule based classifier (image tags)
- Mallet-based classifiers

## Extractor

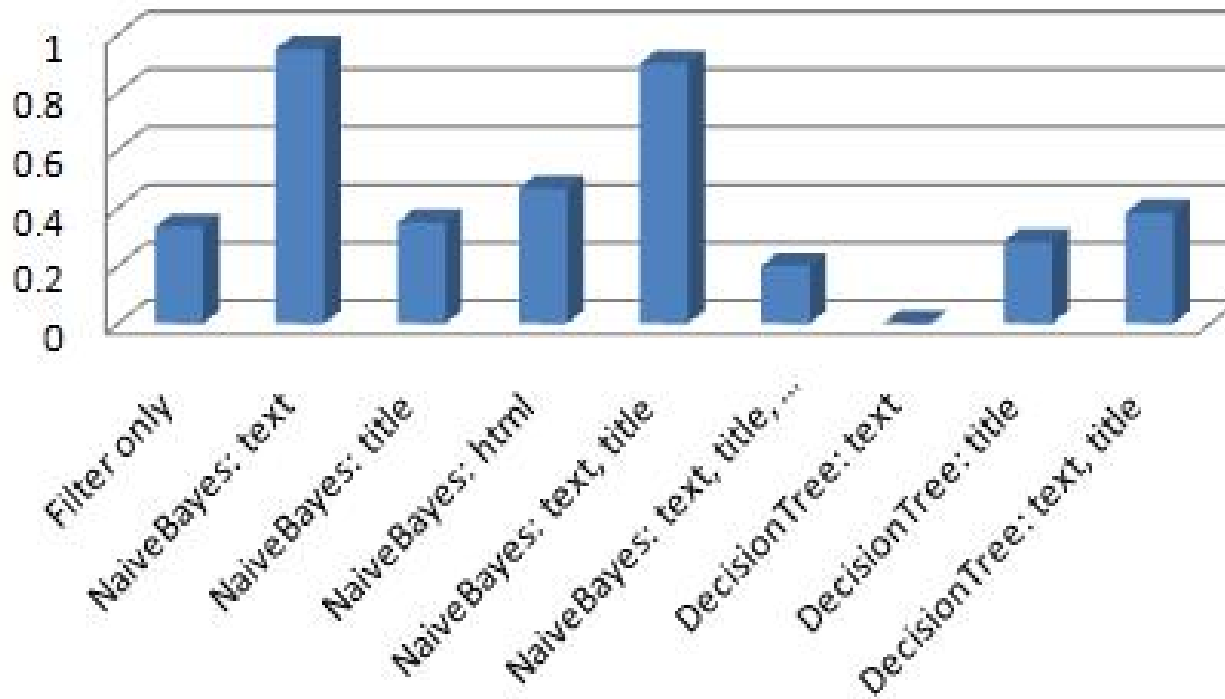
- Location information
- Webcam image, description

## Interface

- [Camcrawler Demo](#)

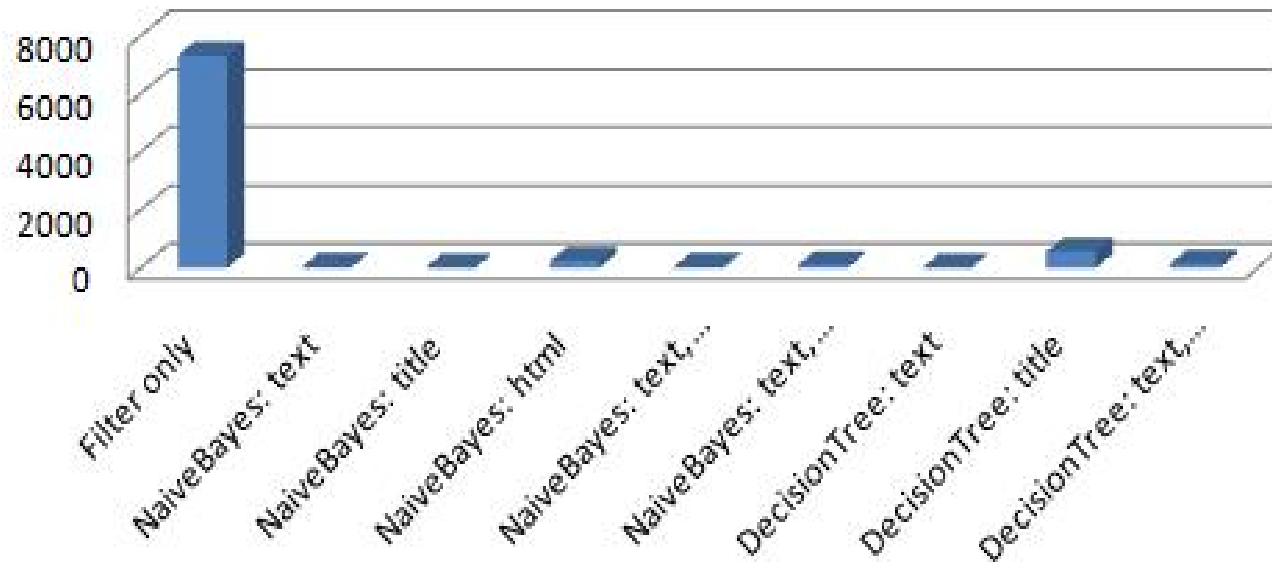
# Evaluation: Classifier

## Classifier Type vs Sample Accuracy



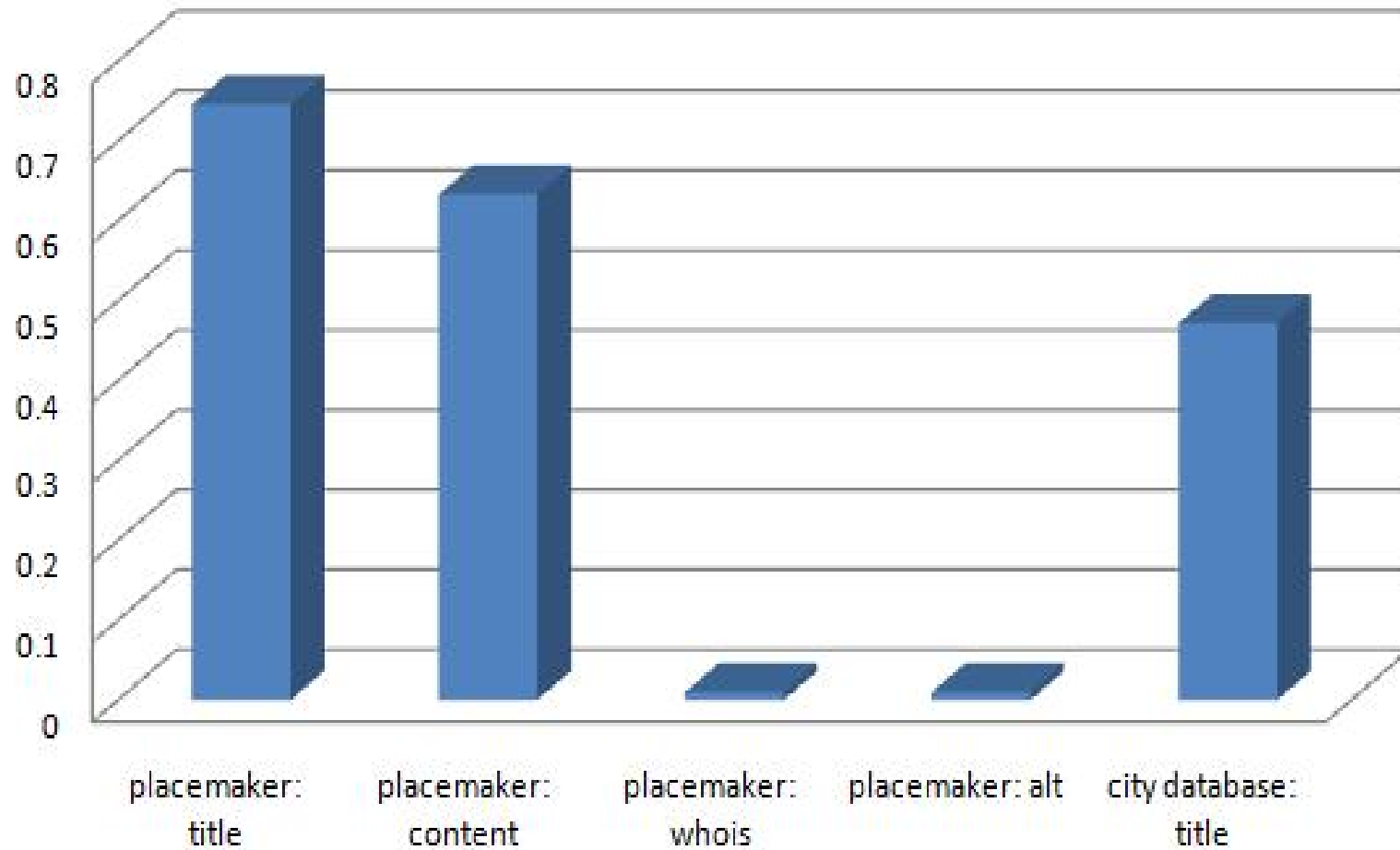
# Evaluation: Classifier

## Classifier Type vs Expected Number of Cams



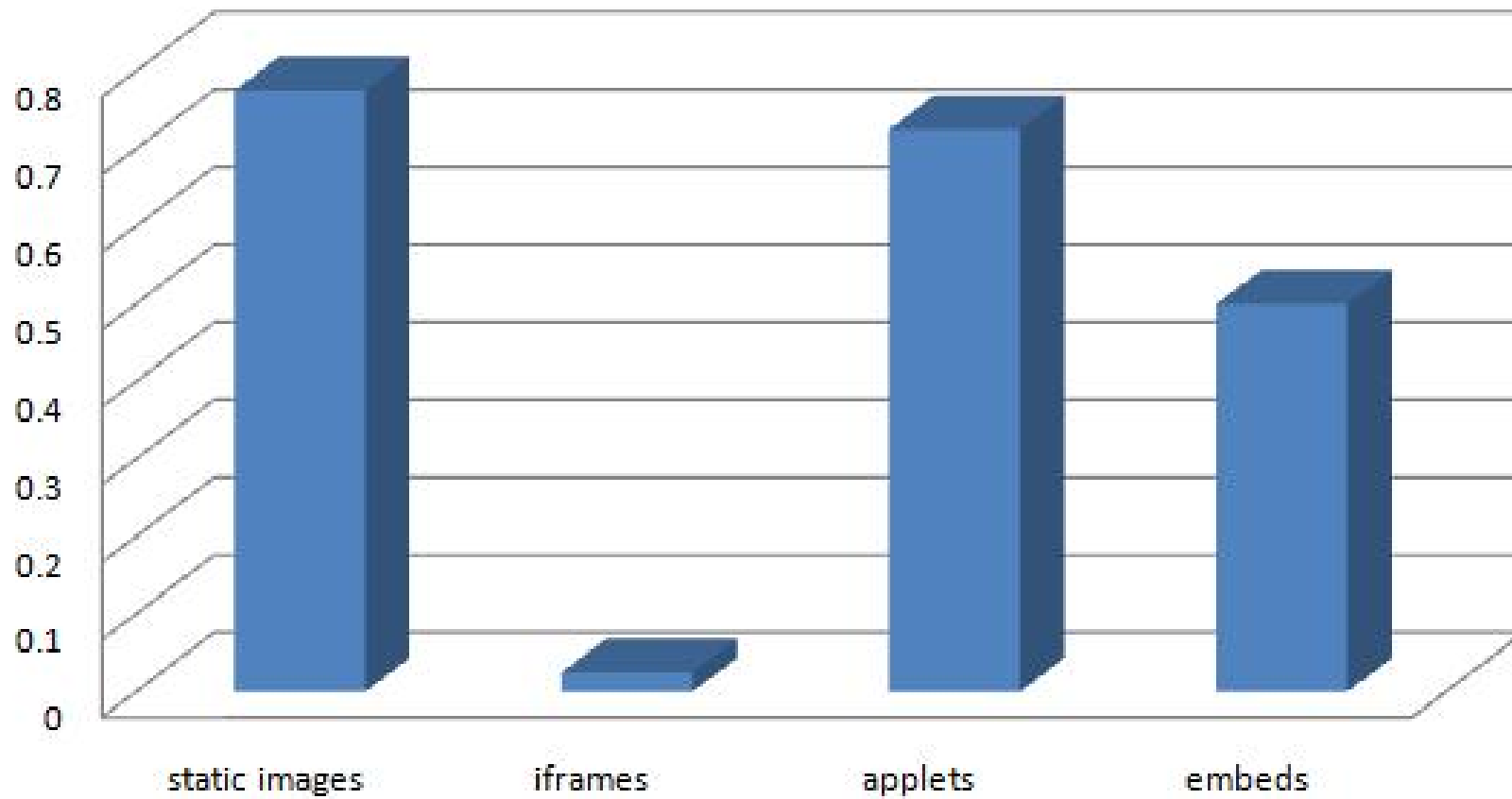
# Evaluation: Location Extractor

## Location Extractor Accuracy



# Evaluation: Webcam Extractor

**Proportion of Correct Extractions  
by Webcam Type**



# Lessons Learned

## Crawler

- Requires lots of drive space
- Easily "wanders off"
- Need much more crawl data for more webcam results

## Classifier

- Keyword based classifiers not helpful

## Extractor

- Location extraction on title is a good classifier
- "whois" proofed to be unreliable
- iframes are unreliable sources of webcam feed



# Division of Labor

**Ricky Bargreen:**

URL Fetcher, Location Extractor, Webcam Extractor, UI

**Flavio Pfaffhauser:**

URL Fetcher, Crawler, Filter, Location Extractor

**Forrest Vines:**

Crawler, Rule-based Classifier, Mallet Classifiers

**Yih Sun Khoo:**

Filter, Rule-based Classifier, Webcam Extractor, UI