



Technical Dimensions of “Responsibility”

**Seventh Search Engines Conference
San Francisco, CA**

**David A. Evans
Clairvoyance Corporation
April 15, 2002**



- **Orientation to the Problem**
- **Content Processing**
- **Information Management (State of IR)**
- **Filtering Based on Classification**
- **Other Types of Filtering**
- **Conclusions**



Where Content Matters

- **Security** (Access Restrictions; Viruses)
- **Copyright** (Blocking Illegal Use, Downloads)
- **Spam** (Blocking Nuisance, Behavior)
- **Pornography** (Offensive Material)
- **Publishing Guidelines** (“No Nazi Memorabilia”)
- **Competitive Intelligence** (Public Discussion)
- **Corporate Policy** (Sexual Harassment; IP; IM)
- **Access to Official Information** (Authority)



The Claims

Filters “Work”

“The government argues that filtering software has vastly improved since the [Children's Internet Protection Act of 2000] law was enacted, making fewer mistakes and allowing libraries to unblock sites that were blocked in error.”—*AP March 25, 2002*

“Justice Department attorney Rupa Bhattacharyya said filters aren't perfect but are efficient enough to support the mandates outlined in the statute, ...”—*AP April 5, 2002*



The Claims

Filters “Work”

“Chris Lemmon of eTesting Labs said tests of four filtering systems found they correctly blocked from nearly 83% to 98% of pornographic sites and incorrectly blocked from none to 7% of unobjectionable sites.”—AP March 29, 2002

“[Lemmon] acknowledged to the judges, however, that when trying to "fool" the filters to tag inoffensive Web sites as *pornography*, testers "stayed away from the gray areas" and used straightforward sites that contained not a hint of racy content.”—AP March 29, 2002



The Claims

Filters “Don’t Work”

“American Library Association lawyer Paul Smith, cited witness testimony that between 6% and 15% of Web sites that are blocked by filtering software are done so incorrectly. Some software can mistakenly brand Web sites on breast cancer, for example, as **pornography .**

... The imperfect technology, combined with the dynamic nature of cyberspace, means filtering software will be "making new mistakes" as fast as the old ones are caught, he added.”—*AP April 5, 2002*



The Claims

Filters “Don’t Work”

Bennett Haselton (www.peacefire.org):

“[We] did a report on pages that were blocked by <<Software₁>>. One of them was a Liza Minnelli fan page, blocked as "adult entertainment." ...

There was a URL for St. Augustine's Confessions that got blocked by <<Software₂>>. ... The [public relations] guy from <<Company₂>> said that their technology was so advanced, their software must have found the page and translated the page from the Latin, and then detected something offensive in the meaning of the translated text. I think it had more to do with the fact that the page had the word "cum" on it so many times, which means "with" in Latin.”—*WSJ April 9, 2002*



Text Content

Child Pornography??

**“Roger,
I’ve just gotten the results of the amnio and
they came back "sex: female"—we’re going to
have a little girl!!”**

**... NOT (... OR “sex” OR ...
... OR (ADJ(“little”, “girl”)) OR ...
...)**



Text Content

Sex Harassment??

**“Roger,
I’ve just been thinking (as I stare into space)
that I’d like to have you YYY my ZZZ.”**

tune ... carburetor

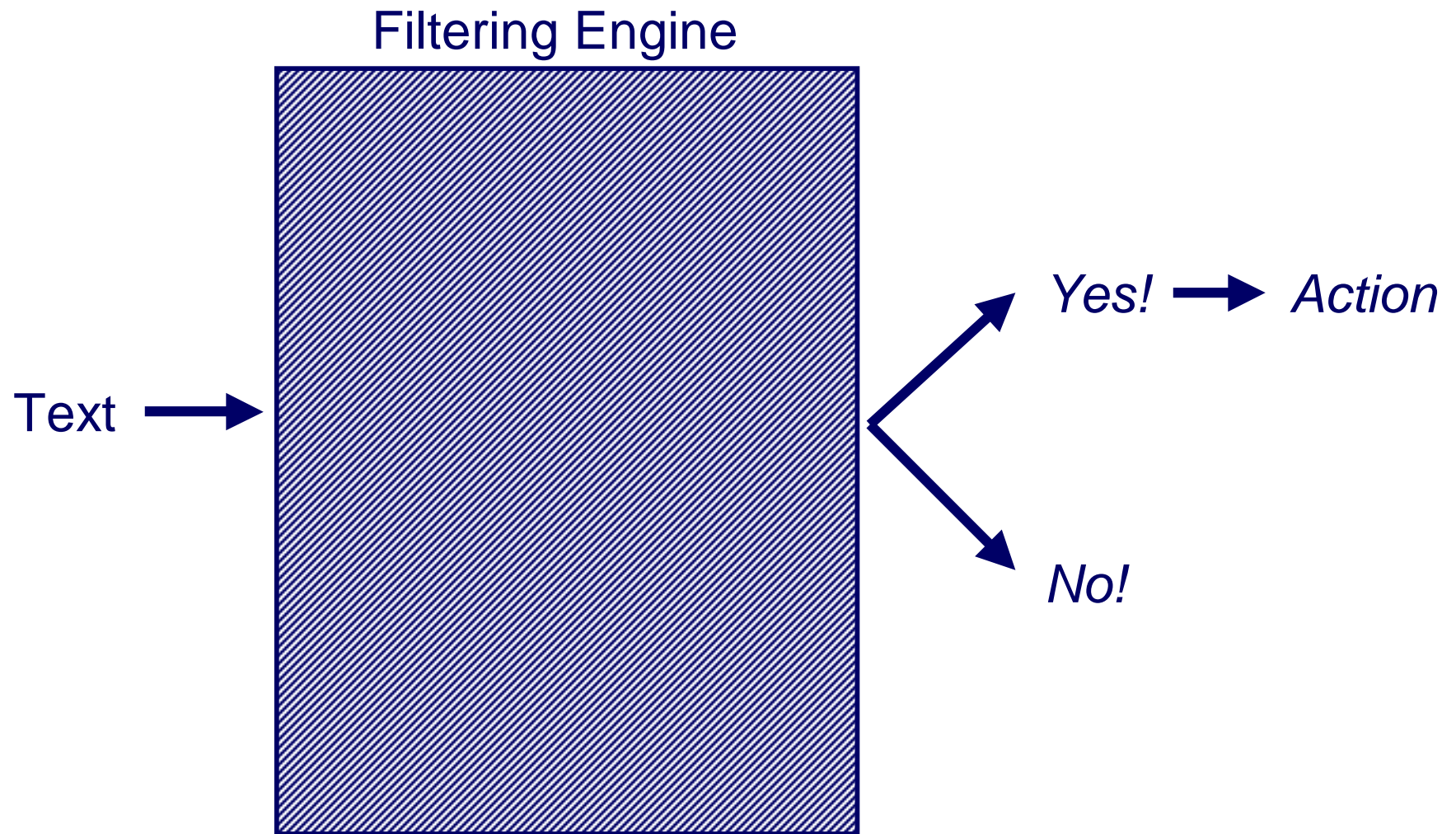
reformat ... hard-drive

baby-sit ... kids

take over ... report



Filtering: General View





Approaches

- **100% Manual**
 - Humans Read the Content
- **Semi-Automatic**
 - Machine Nominates Suspect Content and Human Reviews Decision
 - Human Writes “Rule”; Machine Applies It
- **100% Automatic**
 - Machine Analyzes Content for “Features” and Decides

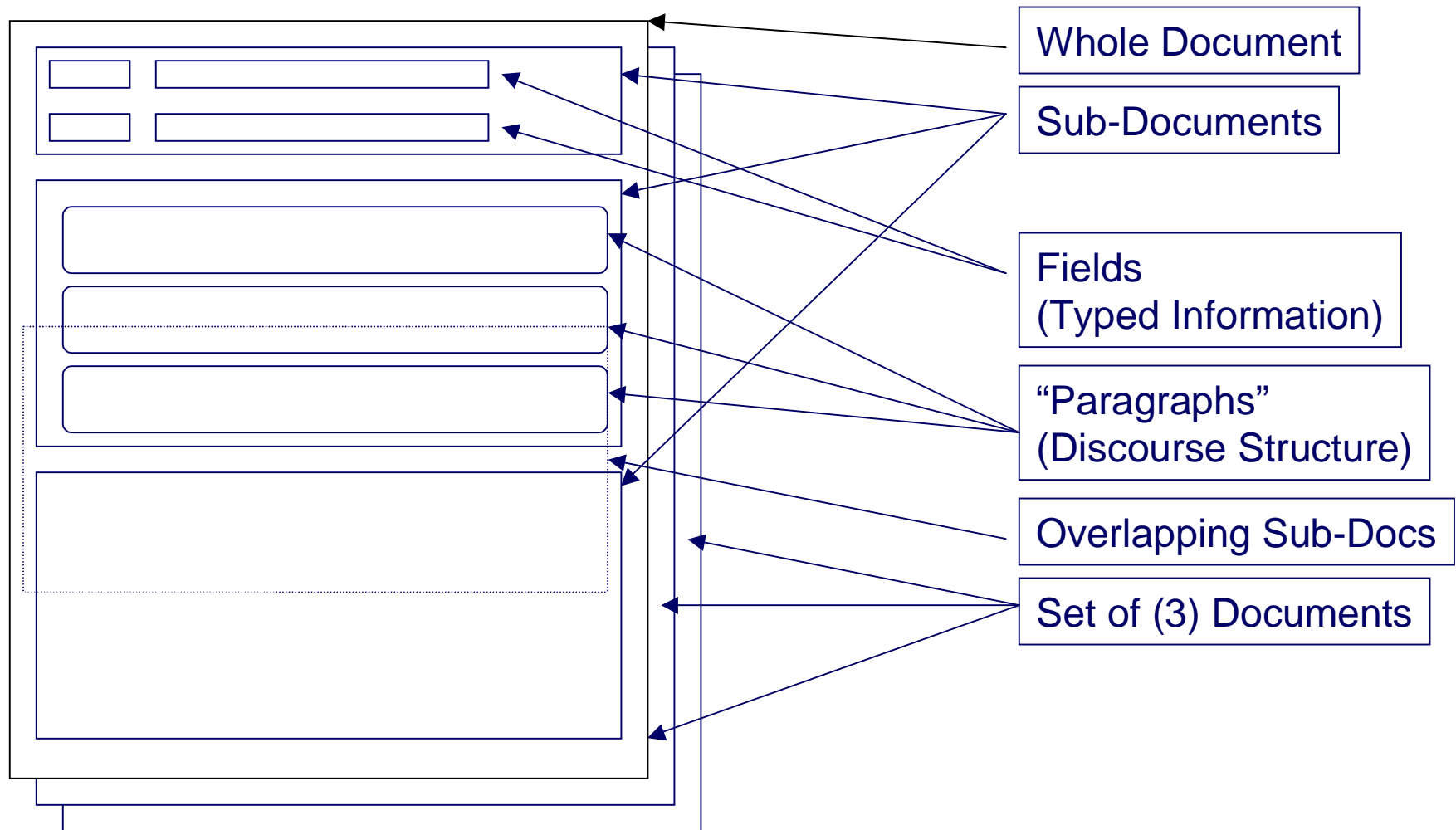


- **Orientation to the Problem**
- **Content Processing**
- **Information Management (State of IR)**
- **Filtering Based on Classification**
- **Other Types of Filtering**
- **Conclusions**



Text Handling

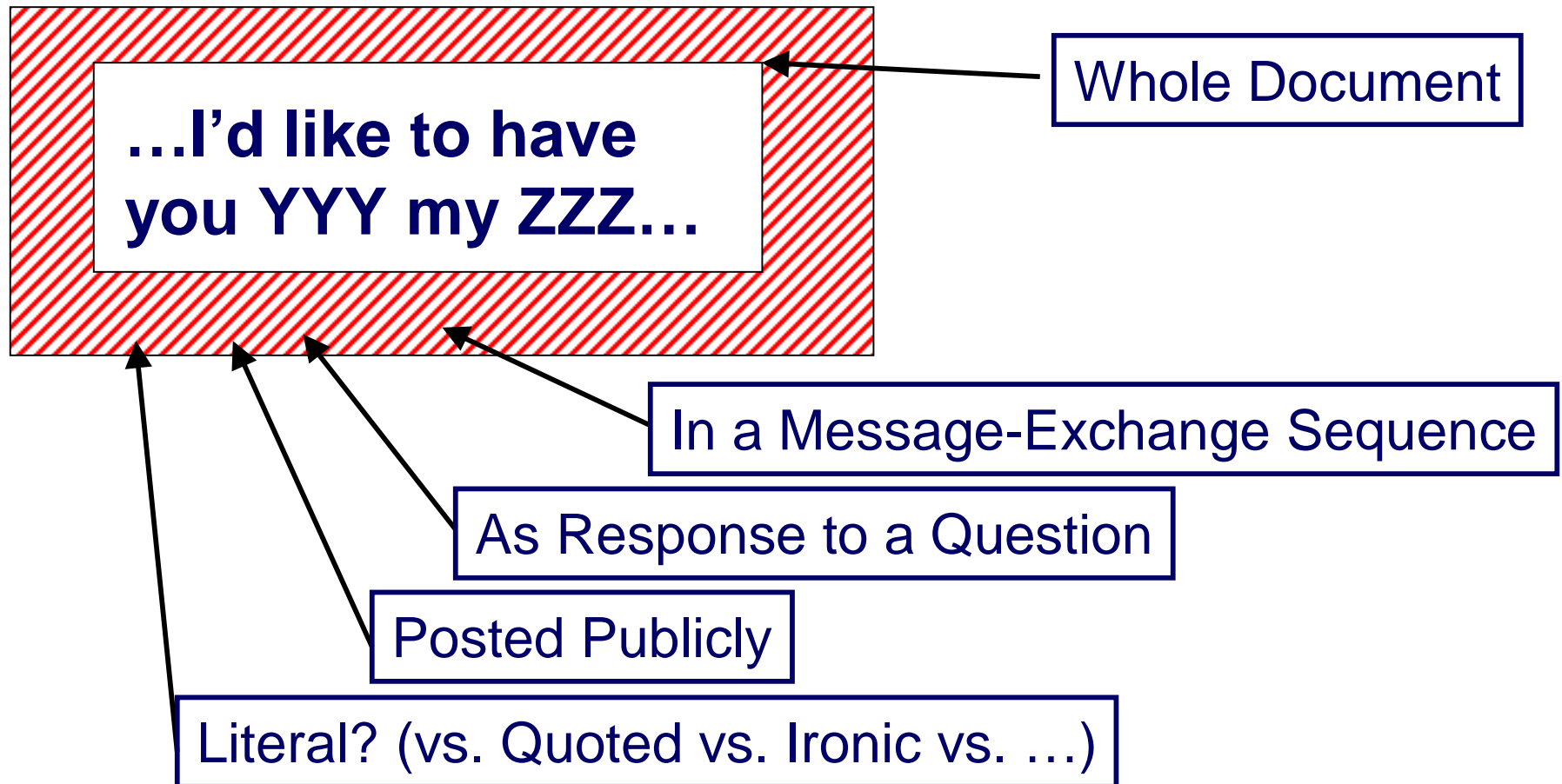
Explicit Context Selection





Text Handling

Implicit Context Selection





Content Representation

Sample Text

“In 1998, The Psychological Bulletin, a highly respected journal, ..., published a review of 59 prior studies of college students who said they had been sexually abused in childhood. The authors concluded that the effects of these encounters were "neither pervasive nor typically intense," ... The authors questioned the practice, common in many studies, of lumping all such cases together as "sexual abuse," suggesting that in some cases they could more accurately be called "adult-child sex" or "adult-adolescent sex." ...”

“Renegade View on Child Sex Causes a Storm”

By Robert F. Worth

New York Times

April 13, 2002



Content Representation

“Words”

The authors questioned the practice, common in many studies, of lumping all such cases together as "sexual abuse," suggesting that in some cases they could more accurately be called "adult-child sex" or "adult-adolescent sex."

abuse	1	common	1	accurately	1
adolescent	1	lumping	1	authors	1
adult	2	practice	1	cases	2
child	1	suggesting	1	practice	1
sex	2	together	1	questioned	1
sexual	1			studies	1



Content Representation

“Tagging”

The authors questioned the practice,

[Det] [N] [V-Past] [Det] [N]
[V] [V-PastPart] [V]

common in many studies, of lumping all such cases

[Adj] [P] [Q] [N] [P] [V] [Pro][Det] [N]
[V] [RelP] [Q] [V]

together as "sexual abuse," suggesting that in

[Adv] [Adv][Adj] [N] [V-Prog] [Det] [P]
[P] [V] [Rel]

some cases they could more accurately be called

[Pro] [N] [Pro] [Mod] [Adv] [Adv] [Cop][V-Past]
[Q] [V] [Q] [V-PastPart]

"adult-child sex" or "adult-adolescent sex."

[N] [N] [N] [Cj] [N] [N] [N]
[Quot1] [Comma] [Quot2] [Period]

97%~99+% Accuracy



Content Representation

“Parsing”

The authors questioned the practice,

[NP]

[VerbPhrase]

common in many studies, of lumping all such cases

[ModPhrase]

[PPhrase]

[PartPhrase]

together as "sexual abuse," suggesting that in

[NP]

[VerbPhrase] [Clause]

some cases they could more accurately be called

[NP]

[NP] [VerbPhrase]

"adult-child sex" or "adult-adolescent sex."

[NP]

[NP]

[NP]

95%~98% Accuracy

Rapid Router - [consumer2\...\Documents]

File Server Tree View Tools Window Help

Source Root: D:\Prototypes\cm2\data

Source Clarit Filter Cluster RDB Filter QueryBE

Stream Source Split Chart Time Chart Responder Msg Mover

40 Documents - 3: [6] 15 MPH Crash Totals My Corolla (Picture Attached) [27.60]

title = 15 MPH Crash Totals My Corolla (Picture Attached)
 AuthorAddr = bcorson@my-deja.com
 Received = 11/13/1999 1:45:32 AM
 Newsgroup = misc.consumers
 Newsgroups = alt.autos.toyota,misc.consumers

.....

Two weeks ago, my 1993 Toyota Corolla DX 4-speed automatic became history on a gridlocked Los Angeles area street. View it at: <http://corson.homepage.com/smashedcar.htm>

I was hit on the right side, mainly near the rear right door, at a speed that couldn't have been more than 15 to 20 miles per hour. The body shop I took it too, which is a preferred body shop with my insurance company and many others, estimated the cost of repairs close to \$8,500.00! The insurance company bought my car for it's market value, which was less than this, rather than even attempt repairs.

Lots of damage. Both doors were involved (the rear door couldn't be shut) bent framing around the doors, bent suspension and floor below, and who knows what else.

Anyway, this closes a chapter in my life owning a Corolla. It had to have several major repairs for the 87,500 miles I owned it (I bought it brand new) including a new transmission at 61,000 miles (1,000 miles out of warranty that Toyota agreed to pay for the transmission, but I had to pay for labor. In all, the car was reliable (aside from the transmission) although I wasn't too happy with it. Just purchased a 2000 VW Jetta as a replacement. Any comments on this crash or my experiences with the Corolla?

Sent via [Deja.com](http://www.deja.com/) <http://www.deja.com/>

Highlighting Options:
 No Terms
 Filtered Terms
 "Terms" tab
 (0)
 No Entities
 Filtered Entities
 All Entities
 3/26-Miner
 Force Selection
 (16)
 Show All Fields
 Mouse-over Entity
 Parse Sel/Doc
 Paragraph Browser
 Words Terms/Ent.
 100 200 0

Dates & Times

Names & Places

Amounts

Specific Issues

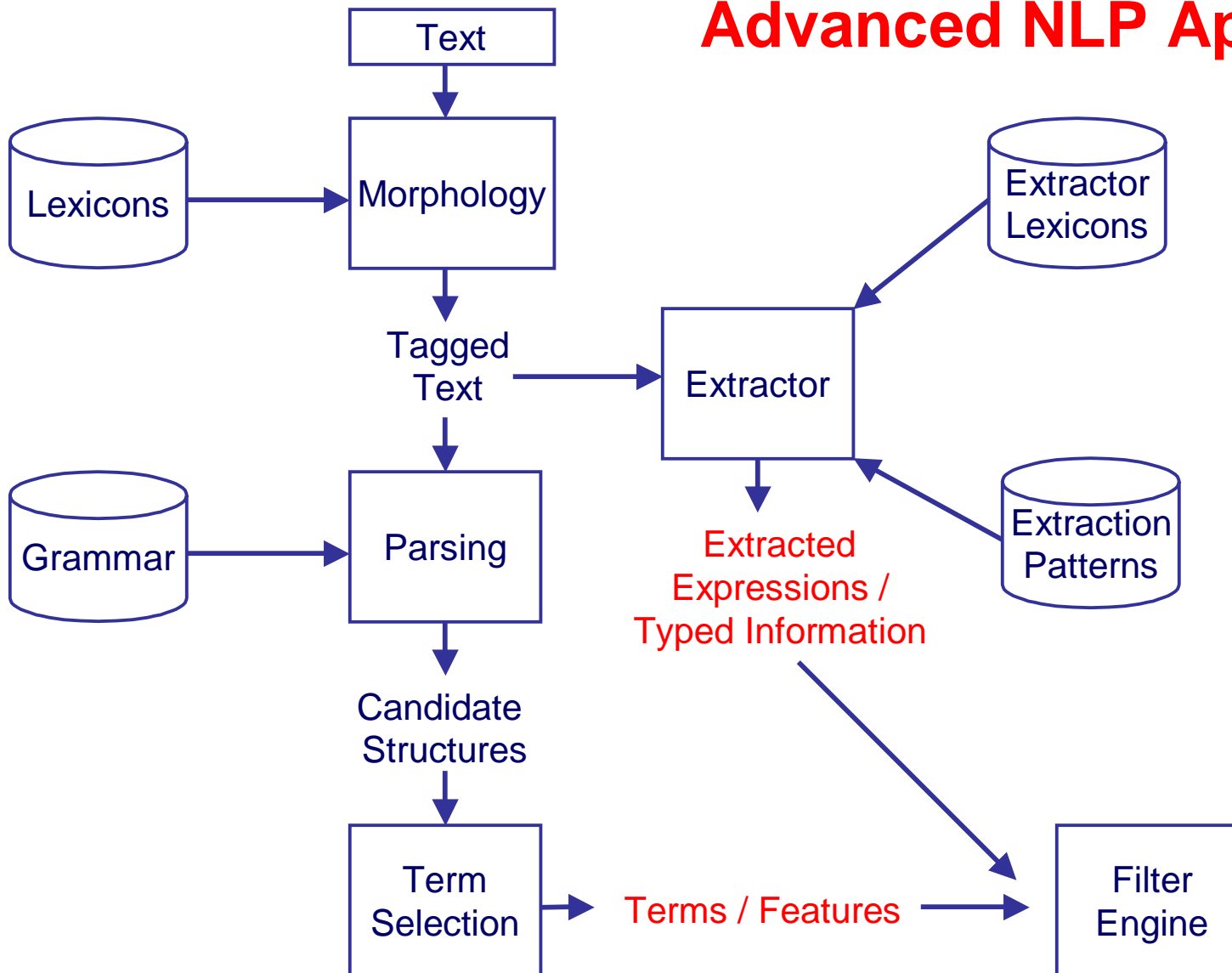
Affect

40%~85% Accuracy



Architecture

Advanced NLP Approach





- **Orientation to the Problem**
- **Content Processing**
- **Information Management (State of IR)**
- **Filtering Based on Classification**
- **Other Types of Filtering**
- **Conclusions**

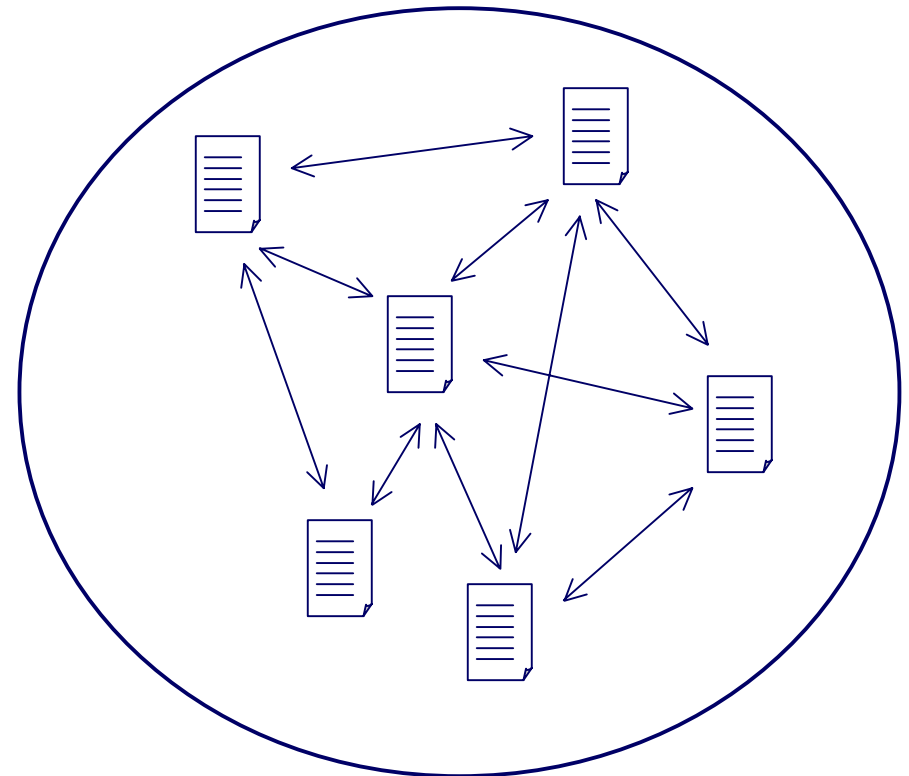


Retrieval Model

Q



*Goal: Optimize
over set of Docs,
return top Docs*



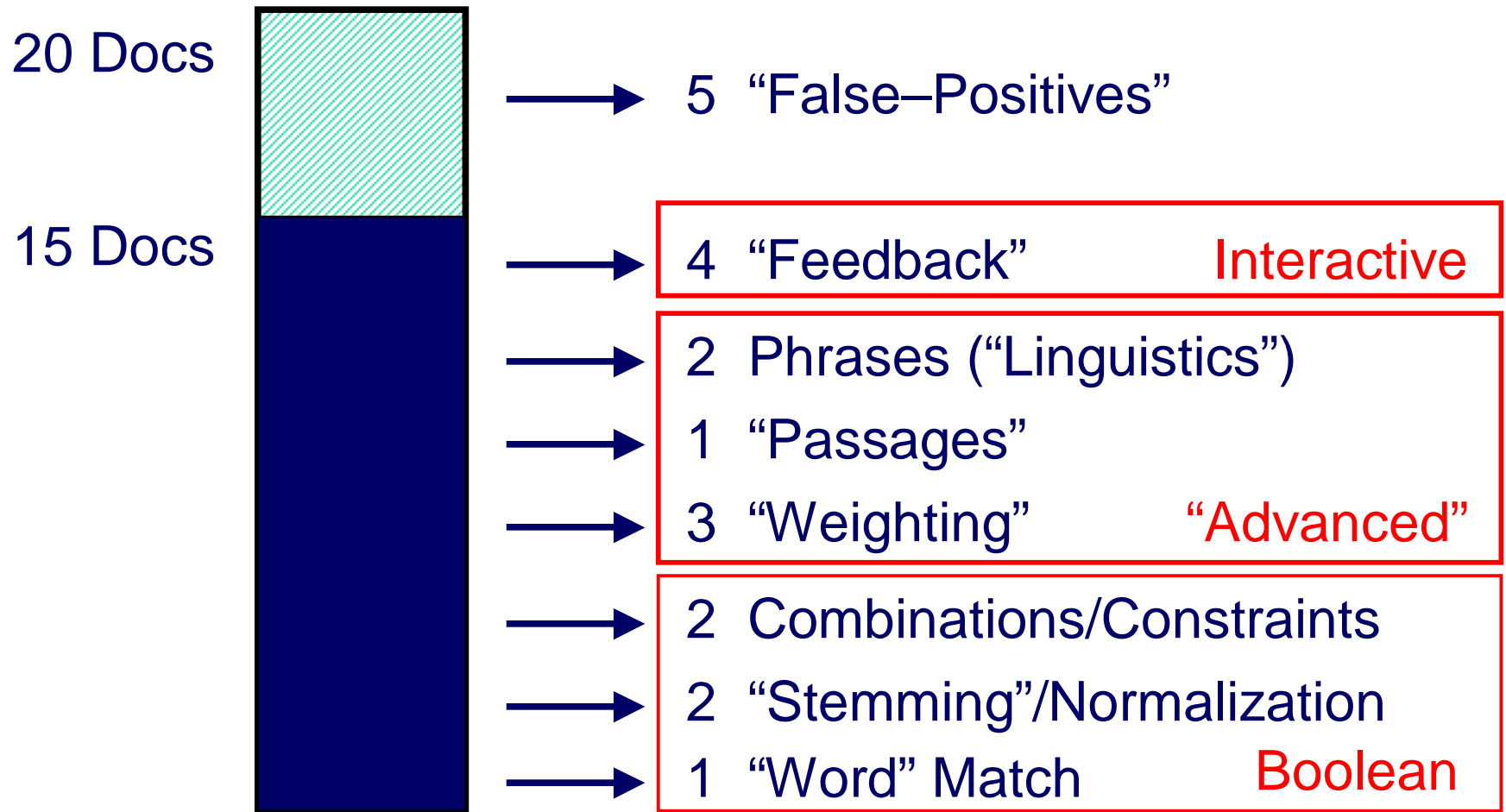
tf

idf



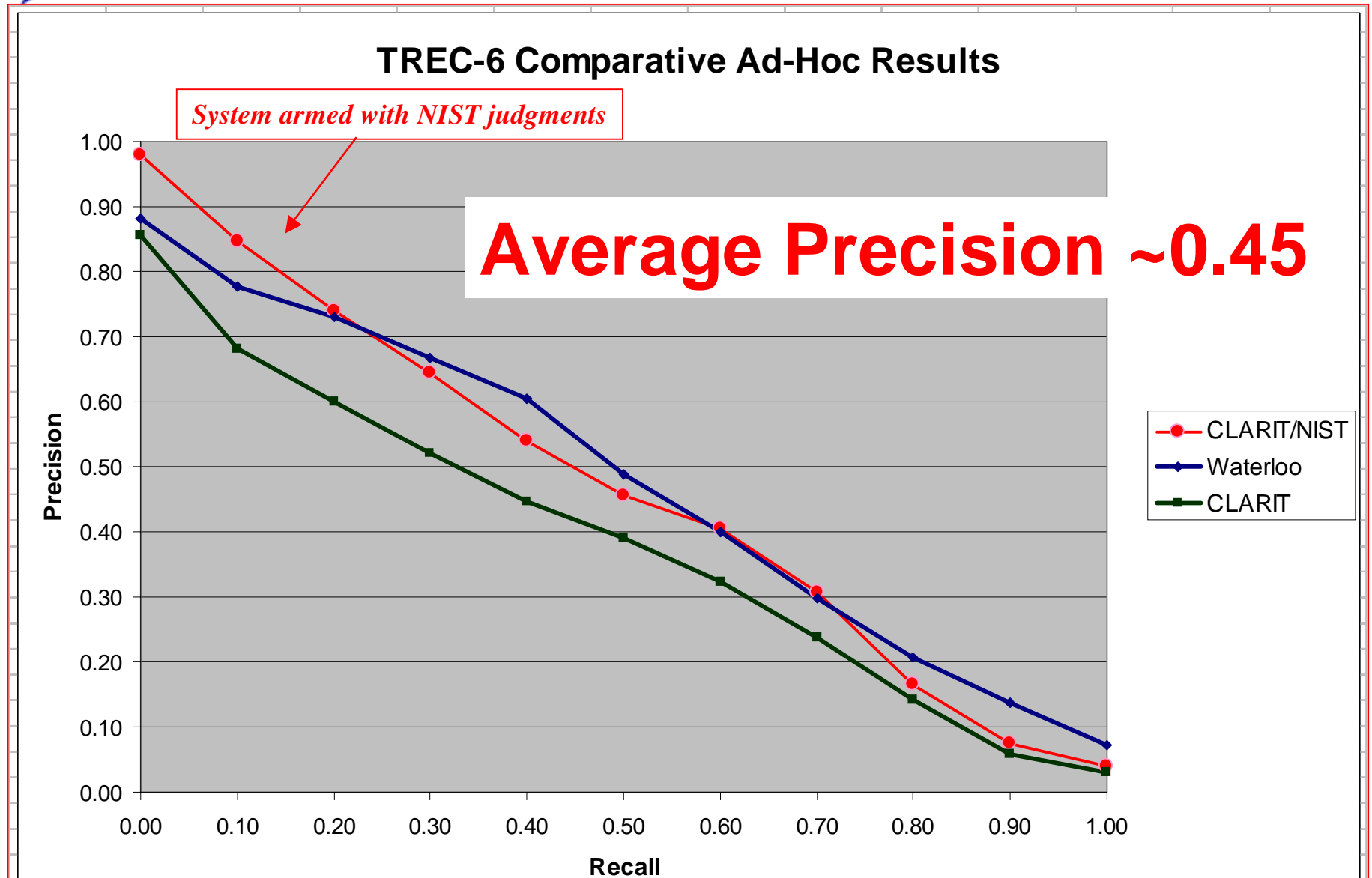
What Contributes to Accuracy?

Good Queries (10+ Terms) / 1M Documents





What is the Limit?





How Much 'Quality' is Possible?

Human Effort vs. Retrieval Performance

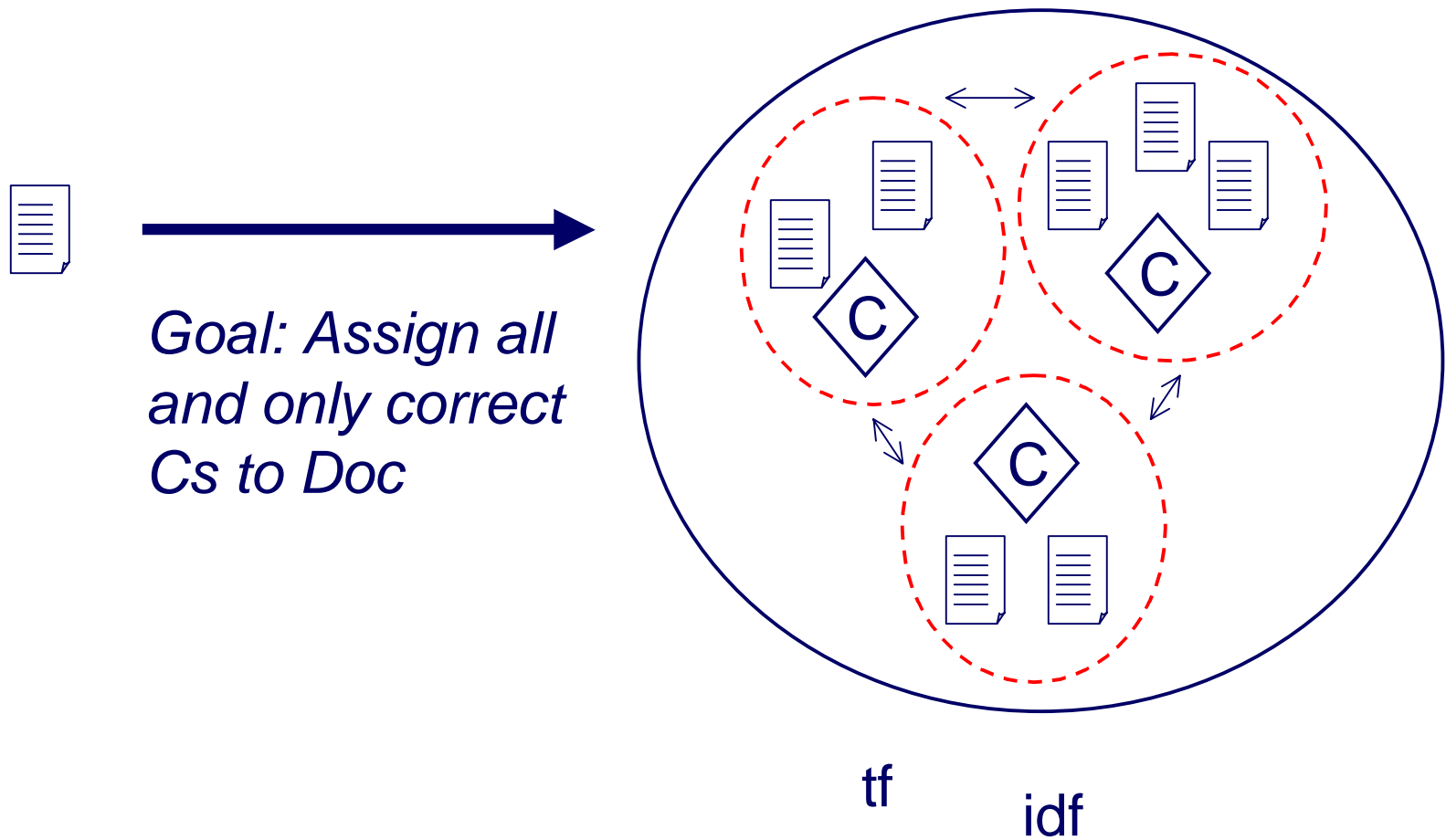




- **Orientation to the Problem**
- **Content Processing**
- **Information Management (State of IR)**
- **Filtering Based on Classification**
- **Other Types of Filtering**
- **Conclusions**



Classification Model





Retrieval Challenge

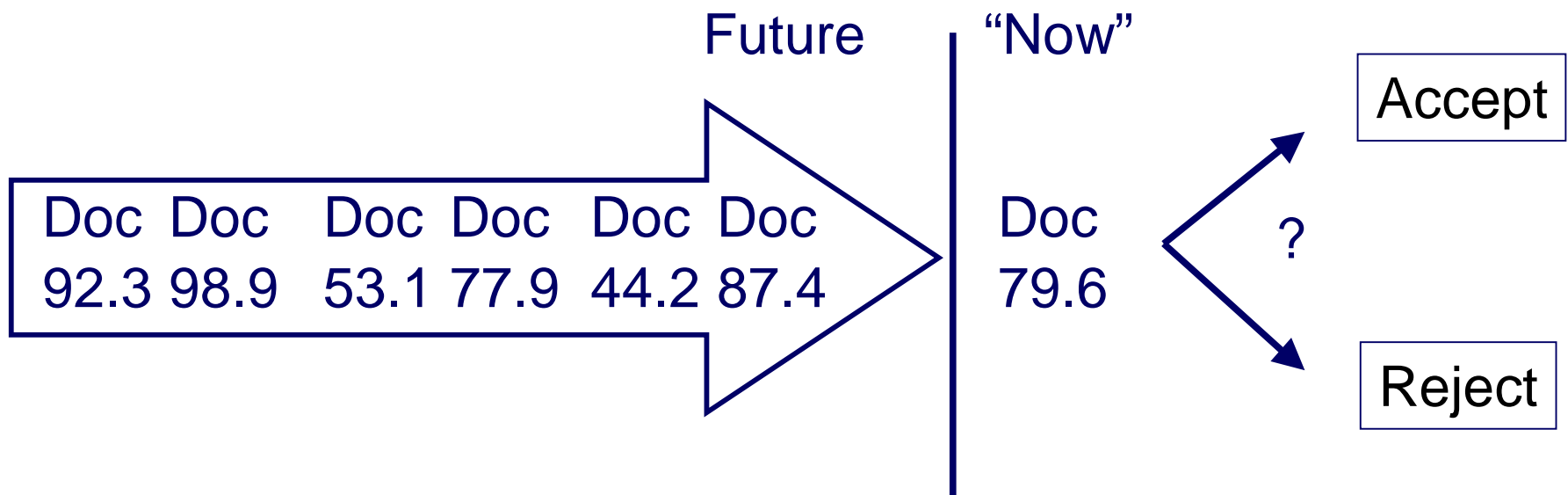
Doc1	98.9
Doc2	92.3
Doc3	87.4
Doc4	79.6
Doc5	77.9
Doc6	53.1
Doc7	44.2



If the process has been optimized, the likelihood of relevance decreases as one goes down the ranked list...



Classification/Filtering Challenge



The process cannot be optimized with respect to the future set of documents...



Elements of the Model

- **Discriminating Features**
- **Thresholds**
- **User Model (Tolerance for Error)**
- **Accommodating “Drift”**



Uses of Examples

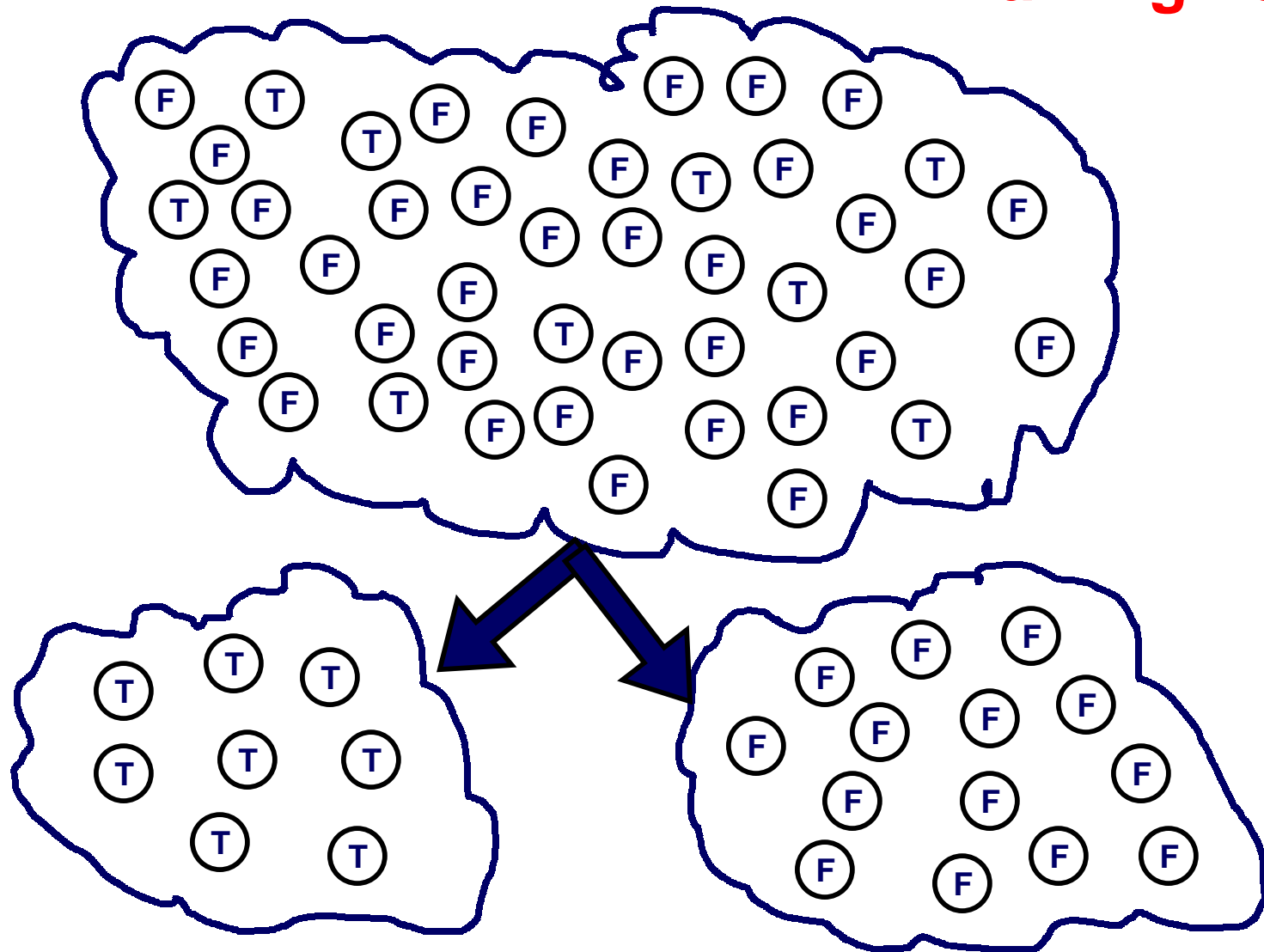
Input to Systems

- **Analysis**
 - Humans → Rules
 - NLP
- **Training**
 - **IR-Based Approaches**
 - Vector-Space Models
 - kNN
 - **ML-Based Approaches**
 - Decision Trees (e.g., C4.5 (Quinlan))
 - Rules (e.g., Ripper (Cohen))
 - Kernel Methods (e.g., SVMs (Vapnik; Joachims))



Uses of Examples

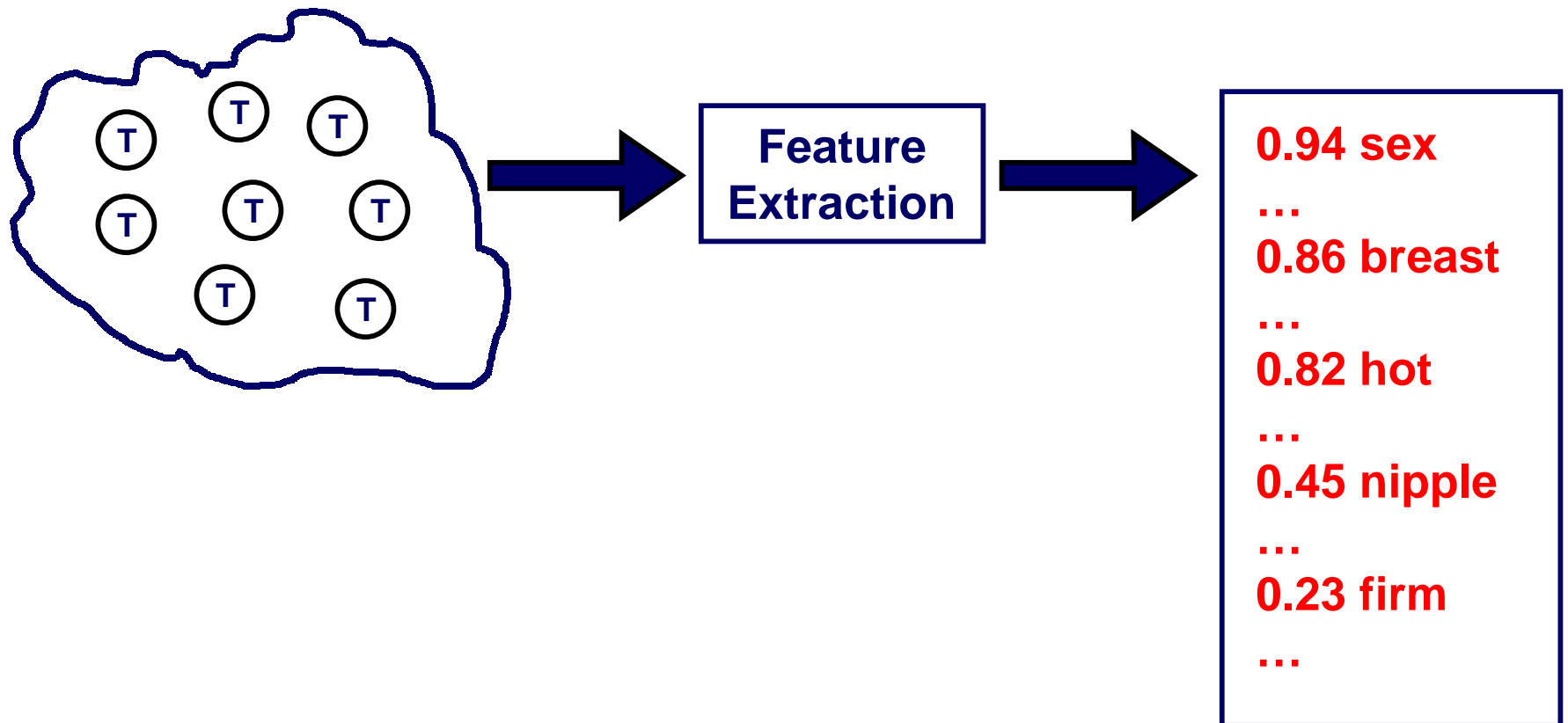
Training Data





Uses of Examples

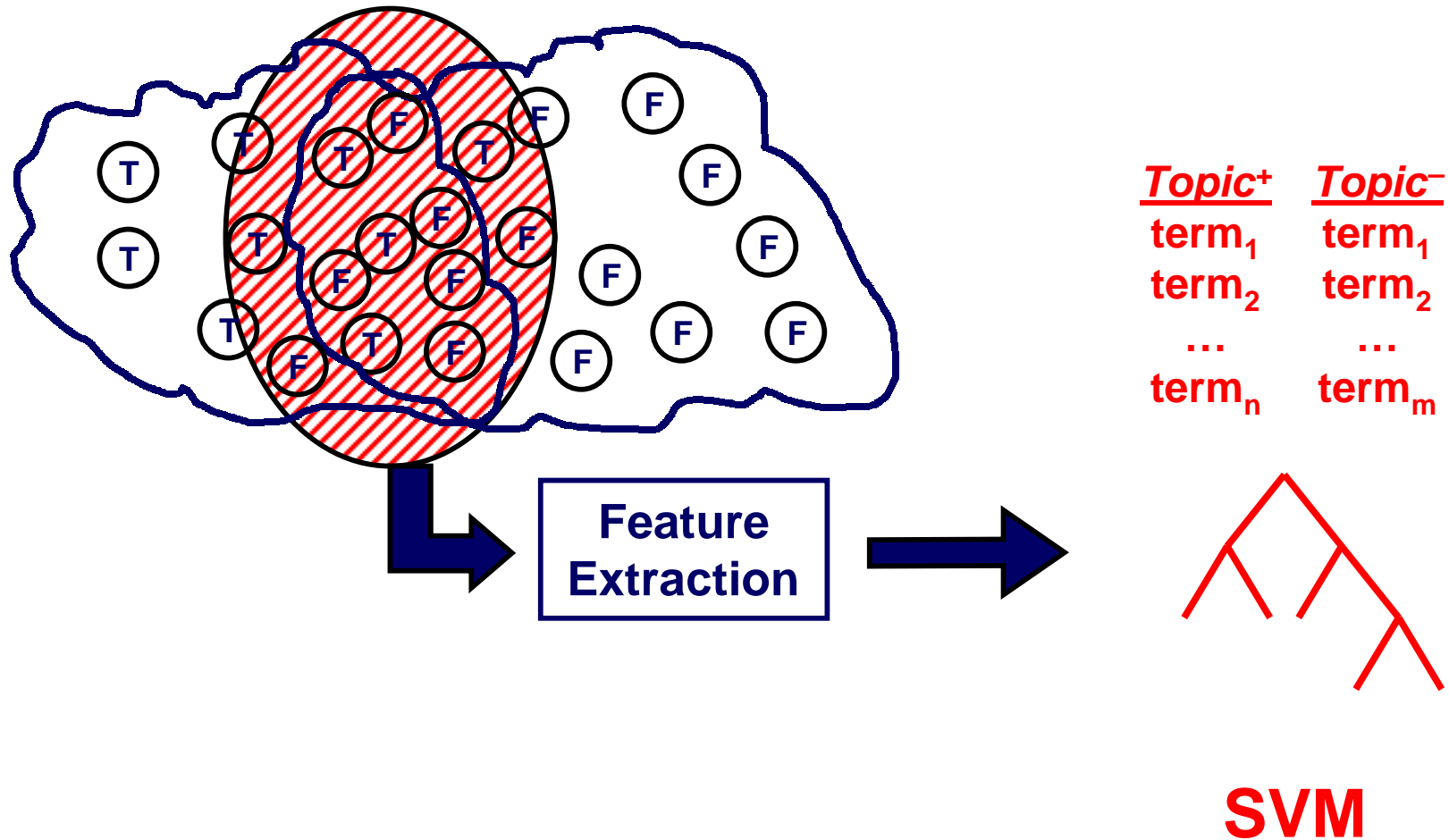
Feature Vectors





Uses of Examples

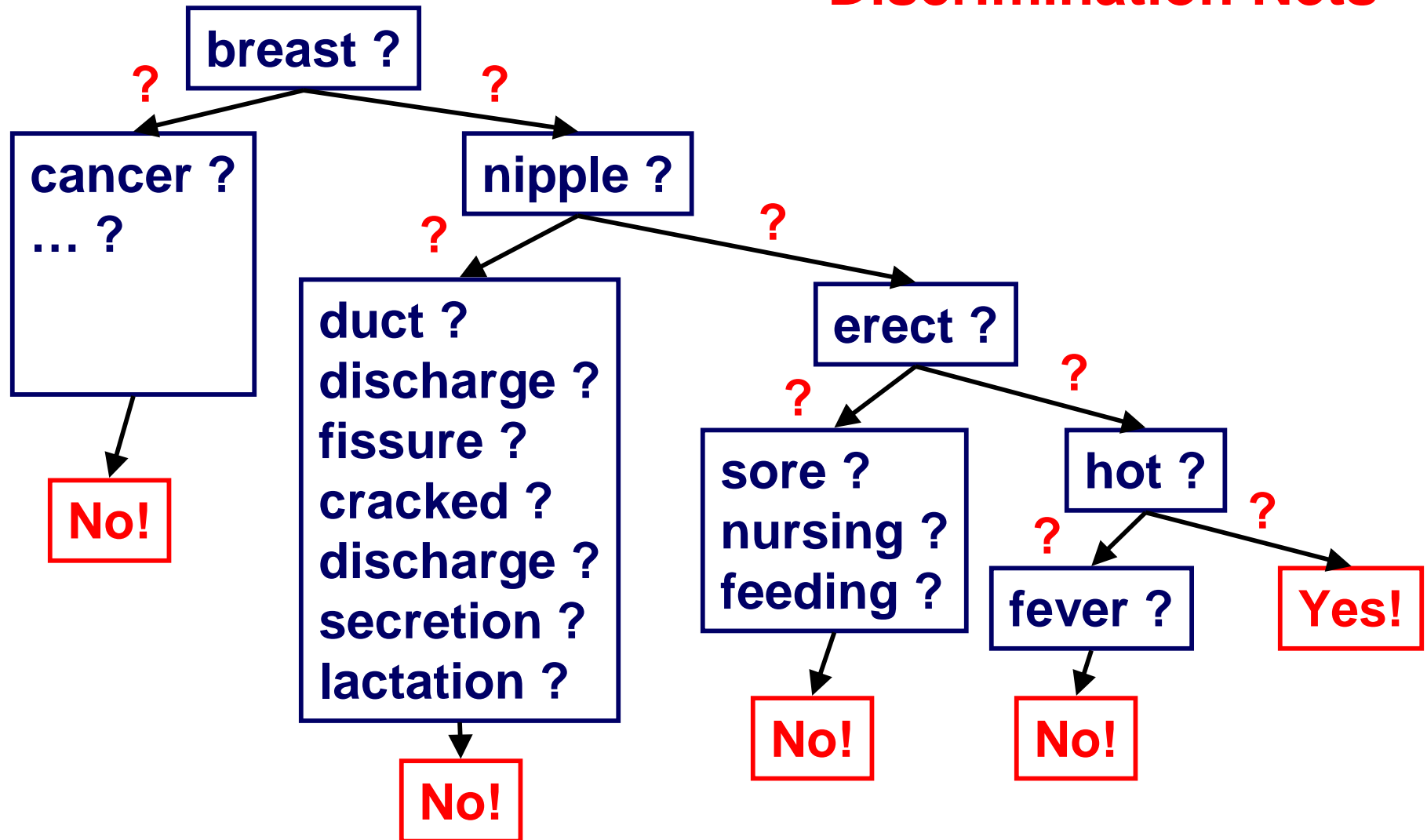
Discrimination, Boundary Conditions





Uses of Examples

Discrimination Nets





Results: Westlaw Subject Cats

Sample Results for Selected Categories [Thompson]

Category	Recall	Precision
Taxation	0.91	0.71
Bankruptcy	0.82	0.69
Product Liability	0.60	0.60
Commercial Law	0.41	0.38
Government Benefits	0.26	0.37

kNN < C4.5 < Ripper



Results: Reuters-21578

Results for 10 most frequent classes [Joachims 1998]

Approach	MicroAvg Breakeven
Polynomial SVM (d=4)	86.0
RBF SVM	86.4
kNN with k=30	82.3
Rocchio	79.9
Decision Trees (C4.5)	79.9
Naïve Bayes	72.0



Results: OHSU-Med

Results for All 23 Categories

Approach	MicroAvg Breakeven
RBF SVM	66.0
Polynomial SVM (d=4)	65.9
kNN with k=30	59.1
Naïve Bayes	57.0
Rocchio	56.6
Decision Trees (C4.5)	50.0

Results: TREC 2001 Batch Filtering

Results for All 84 Categories

Approach	T10SU
SVM [Lewis 2001]	~0.41
kNN [Ault et al. 2001]	~0.30
RBF SVM [Mayfield 2001]	~0.28
IR [KUN 2001]	~0.26

LESSON

Polynomial Kernels performed very well but required **three (3) weeks of training, whereas most of the IR-based approaches took **three (3) hours or less****



- **Orientation to the Problem**
- **Content Processing**
- **Information Management (State of IR)**
- **Filtering Based on Classification**
- **Other Types of Filtering**
- **Conclusions**



Filtering on “Events”

Topic Detection

- **Category does not Exist “In Advance”**
- **Category may have General Characteristics, but is not Defined by Content (Alone)**
- **Must be Detected, Modeled (Possibly without Confirmation or Human Intervention)**
- **Examples: Spam, Hate Mail, ...**

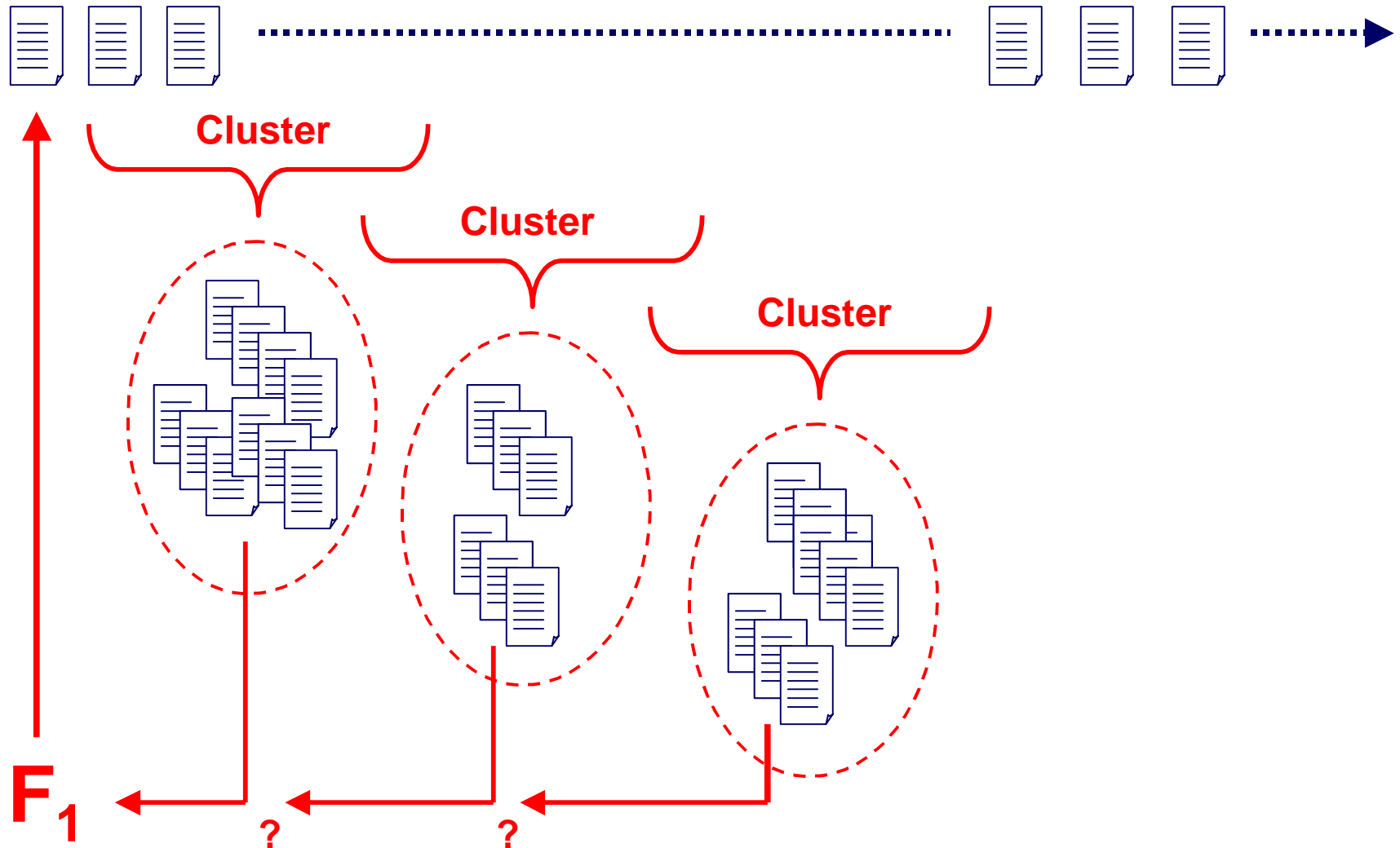


Example: Spam

- **New Spam is Constantly Created**
- **Only Loosely Modeled by Content (Especially in Isolation)**
- **Highly Confusable (cf. “Job Applications”)**
- **Obvious Features are Often Obscured**
- **“Behavioral” Signature (for SPs)**
 - “Attack” can Last 24–72 Hours
 - Can Consume 60% of Total Capacity



TDT Problem





- **Orientation to the Problem**
- **Content Processing**
- **Information Management (State of IR)**
- **Filtering Based on Classification**
- **Other Types of Filtering**
- **Conclusions**



The Process

With Respect to a Particular Method

- **Discover the Topic (Collect Data)**
- **Represent It (Extract Features)**
- **Optimize It (Train on Examples)**
- **Apply It (Run on New/Real Data)**
- **Maintain/Improve It (Use Feedback)**



- **Filtering is a Challenging Problem**
 - Content Processing is Good for Lower-Level Features, More Problematic for Abstractions (Semantics & Pragmatics)
 - The Best IR Results (on Ideal Data)—with no Training—are Far from Perfect (but may be near Practical Limits)
 - The Best Classification Results (on Ideal Data)—with Significant Training are Not Perfect
 - Many Types of Filtering—more Difficult than Classification Tasks—remain to be Mastered
- **Be Cautious With Claims!**



The End



On to the Panel ...