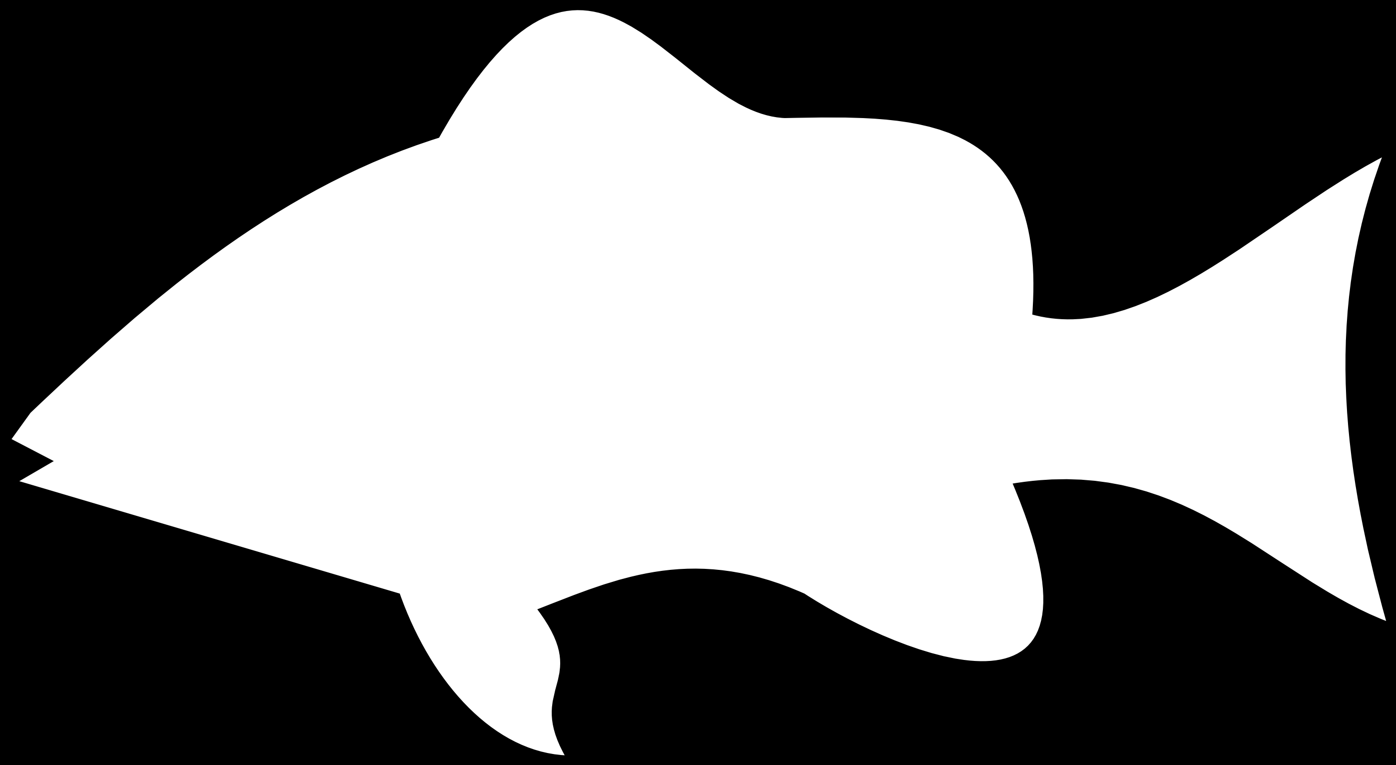
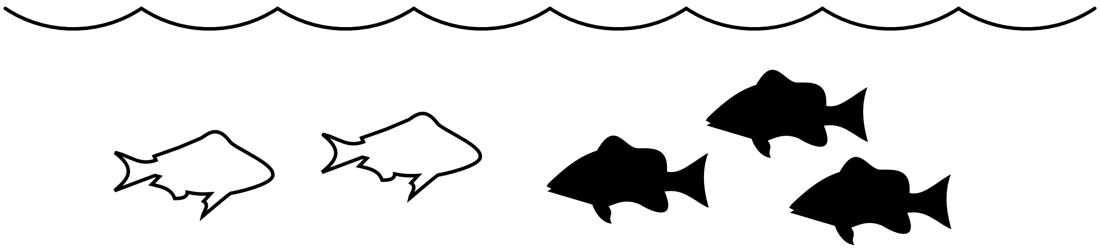


Search Metrics

James Keuning
version 1

Explaining search metrics using fish.



Categories of search results are hard to explain. They are not hard to *understand*, just hard to explain. Anyone who has used the internet to search for anything realizes that of the millions of search results, some meet the needs of the searcher, while others do not. Furthermore, of the billions of items which were searched, some were delivered in the results, some were not. Taken even further yet, of the items which were not delivered, some meet the needs of the searcher, while others do not.

> Results that meet the needs of the searcher are **Responsive**.
> Results that do not meet the needs are **Nonresponsive**.

This marks the difficult (yet, not difficult) point: the search results, those things that are presented to the searcher, indicate the way that the search protocol responded to the query. It could be said that those things (the search results) are *responsive* to the *search*. In other words, after running a search, we might ask the question:

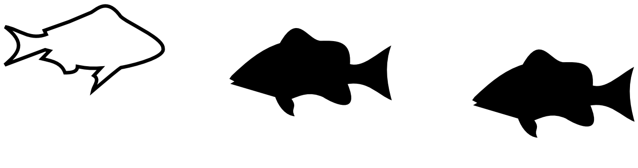
aka "hits"

"could" be said,
not "should"

"How many **responsive** items are there?" — remember

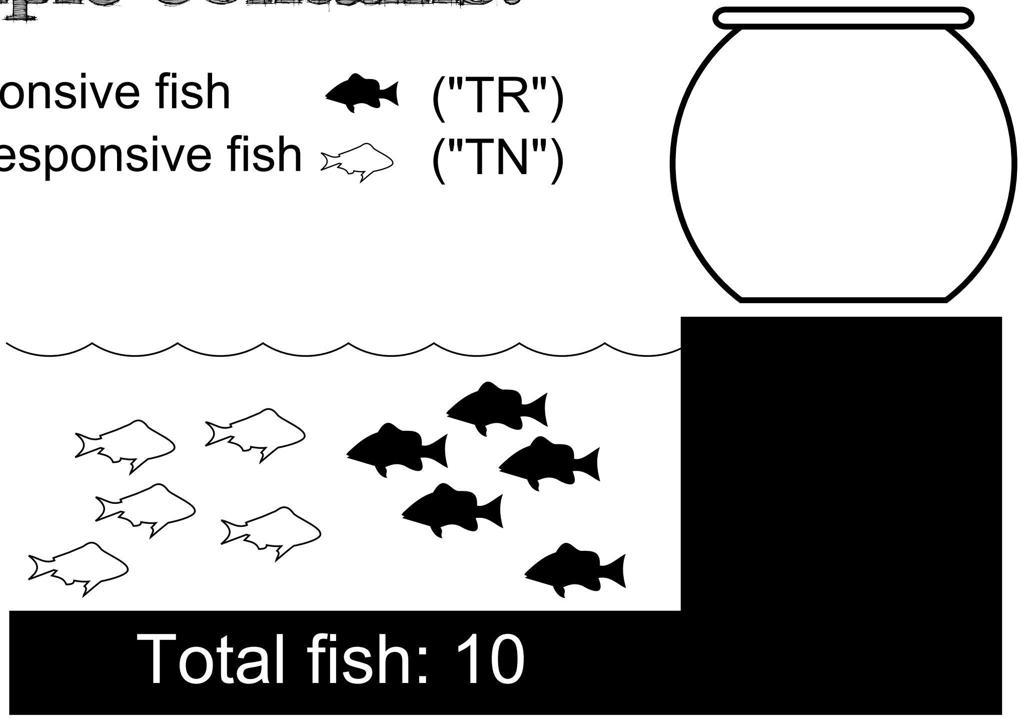
Do not answer this question by providing the number of items in the search results; some of the results may not meet the needs of the searcher and are thus **nonresponsive**.

Items are **responsive** irrespective of how the search handles them. In fact, we measure the effectiveness of the search based on how **responsive** documents are handled. And right now we are going to use fish to explain this.



Example Contains:

5 Responsive fish ● ("TR")
5 Nonresponsive fish ○ ("TN")

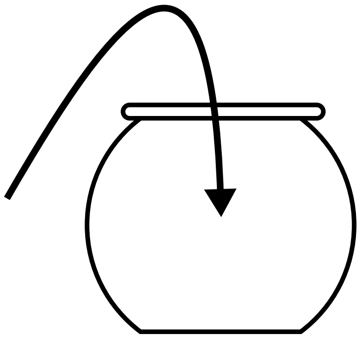


TR and TN: True Responsive and True Nonresponsive, respectively. These are counts before any type of query is run. This "true" number is a bit of a fiction because the responsiveness of a document is subjective. These numbers represent the the result of a 100% perfect search.

see example, below, for subjectiveness

Now we run a search. The fish which the ***search* deems responsive** are put into the bowl. Keep in mind that the search will not be 100% accurate so some nonresponsive fish might end up in the bowl and some responsive fish might get left behind.

Search Results

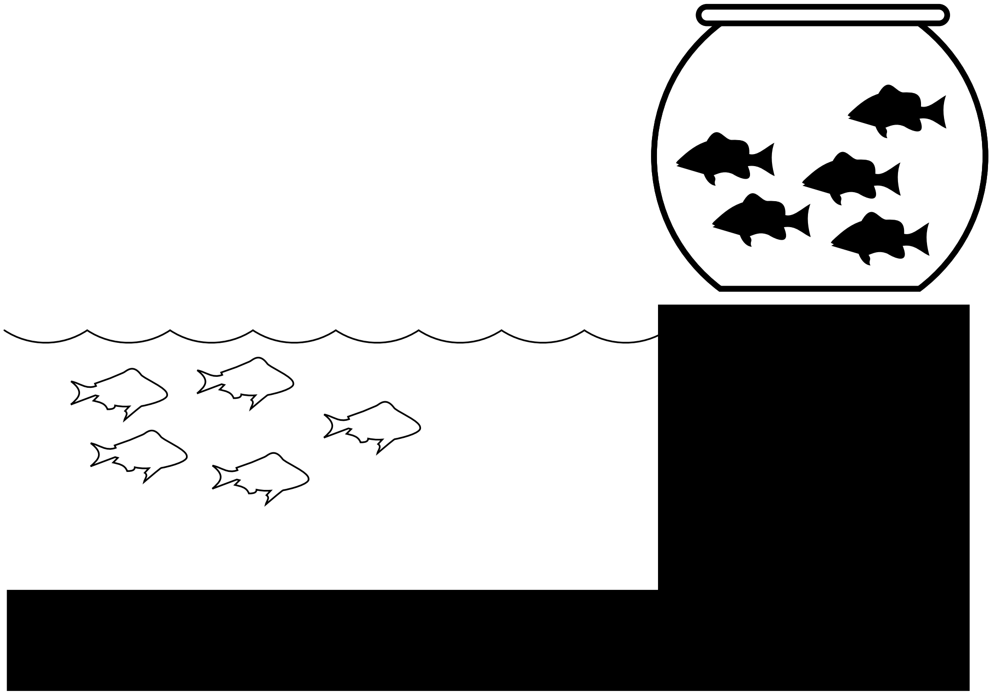


If I run a horrible search and get zero fish in the bowl, it does not change the fact that there are five responsive fish.

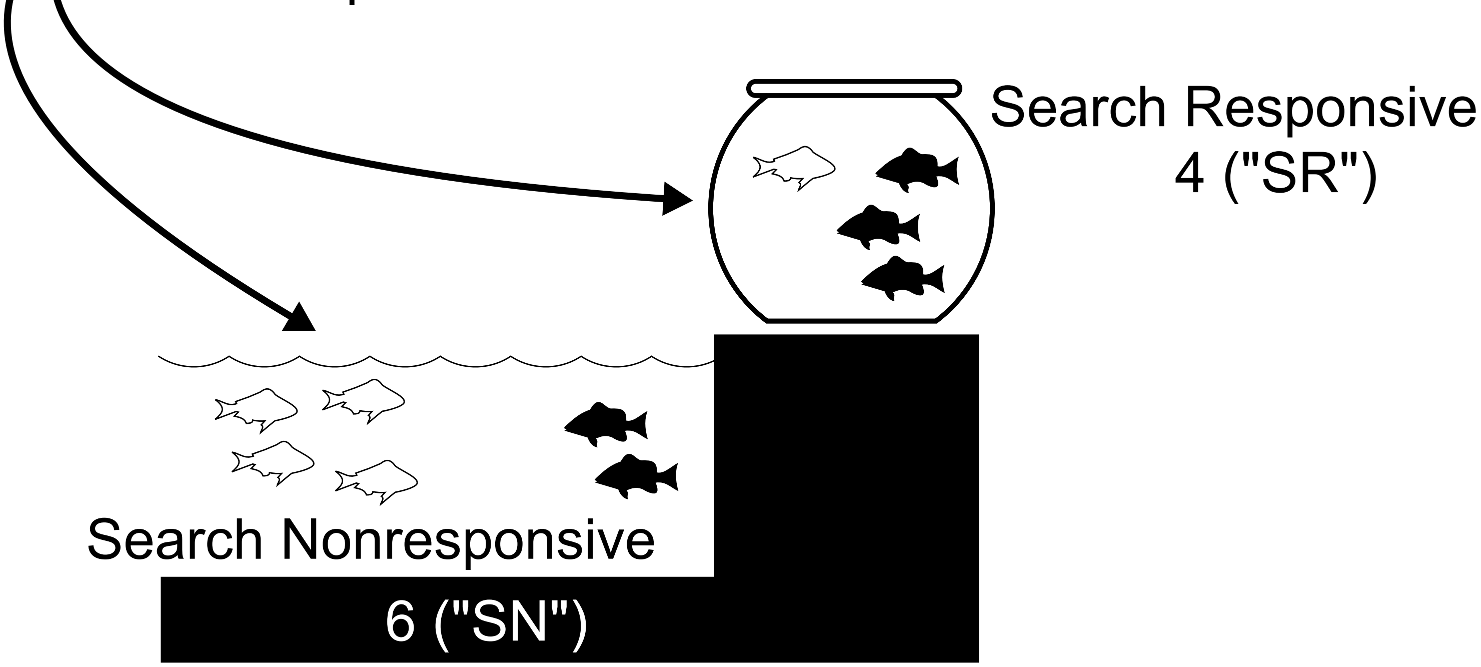
The bowl will contain the documents that the *search calls responsive*.

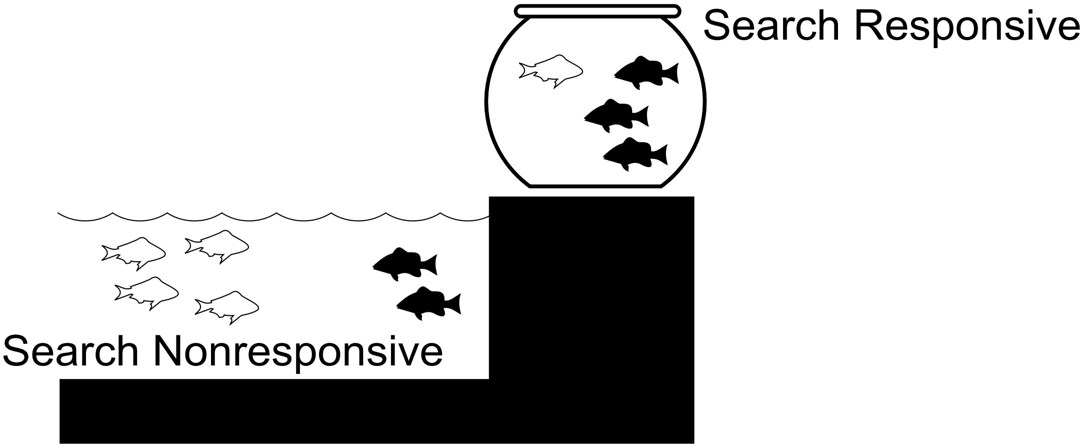
We will call these "search responsive"

This is what perfection would look like.



Perfection rarely happens.
These results are more realistic and will
make sample calculations more useful:





True Responsive	5	TR
True Nonesponsive	5	TN
Search Responsive	4	SR
Search Nonresponsive	6	SN

These numbers did not change

●	fish in the bowl
○	fish in the water

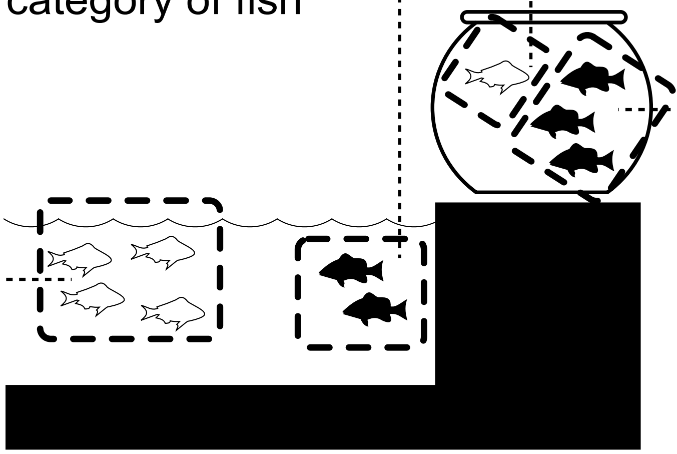
Drop these numbers around the Contingency Table, aka Confusion Matrix*

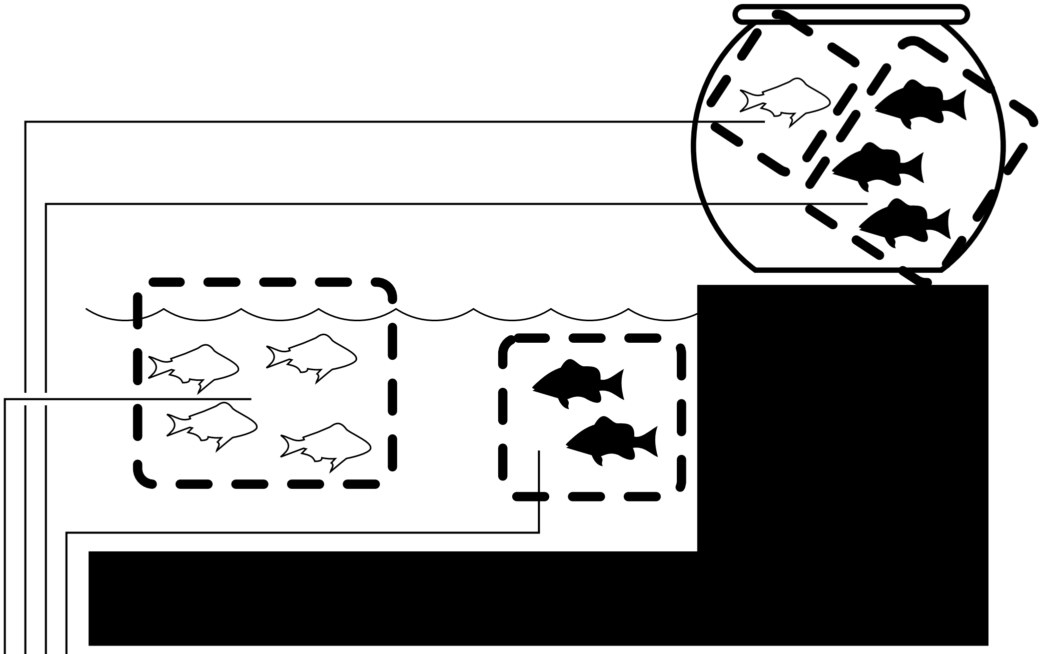
Dichotomous binary classification - yes or no. *No ranking*

	SR	SN	
TR	3	2	5
TN	1	4	5
	4	6	

Notice about this table:
The ROWS represent TRUE values
The COLUMNS are SEARCH values

And take a look at the interior numbers: 3, 2, 1, and 4 correspond to a category of fish





The post-search fish categories have names:

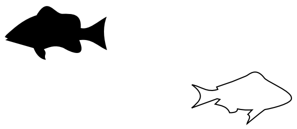
Correct Positive (CP)	correctly marked responsive	black fish in the bowl
False Positive (FP)	incorrectly marked responsive	white fish in the bowl
False Negative (FN)	incorrectly marked nonresponsive	black fish in the water
Correct Negative (CN)	correctly marked nonresponsive	white fish in the water

Overlay these categories onto the table:

	SR	SN	
TR	CP 3	FN 2	5
TN	FP 1	CN 4	5
	4	6	

Stop here and clear the air around this True vs Search responsiveness question.

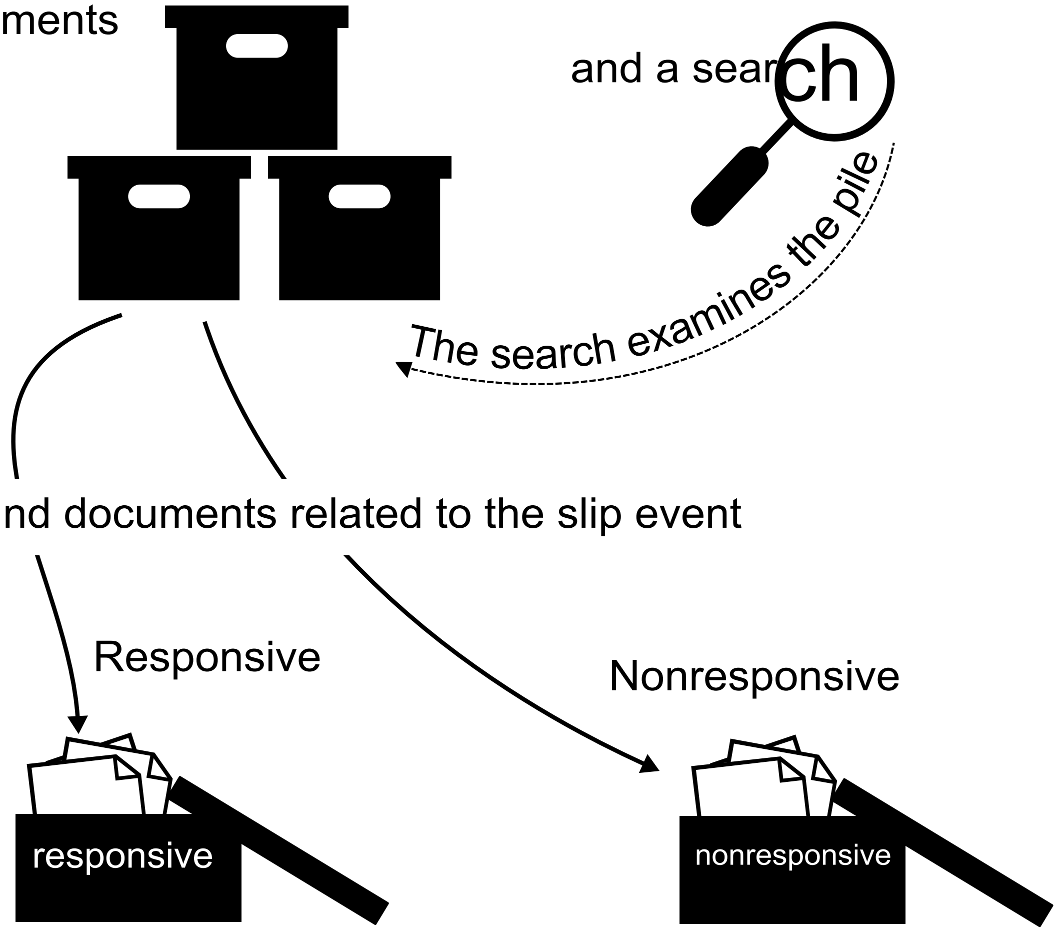
When we are dealing with black and white fish.
The responsiveness question is easy.
It's black and white. Literally.



In real life, it is not so easy.

Imagine a slip and fall case.

and a pile of documents



and a search

The search examines the pile

The Goal:

Find documents related to the slip event

Responsive

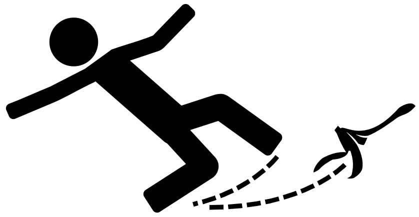
Nonresponsive

POP QUIZ

The search calls the following Responsive. Categorize them.

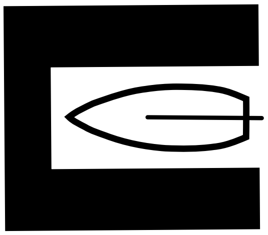
He slipped on the floor.

CP FP FN CN



I reserved a slip at the marina.

CP FP FN CN



It slipped out.

CP FP FN CN

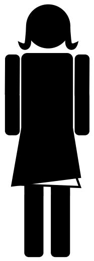
The search finds the following Nonresponsive. Categorize them.

Dude busted his melon.

CP FP FN CN

Her slip is showing.

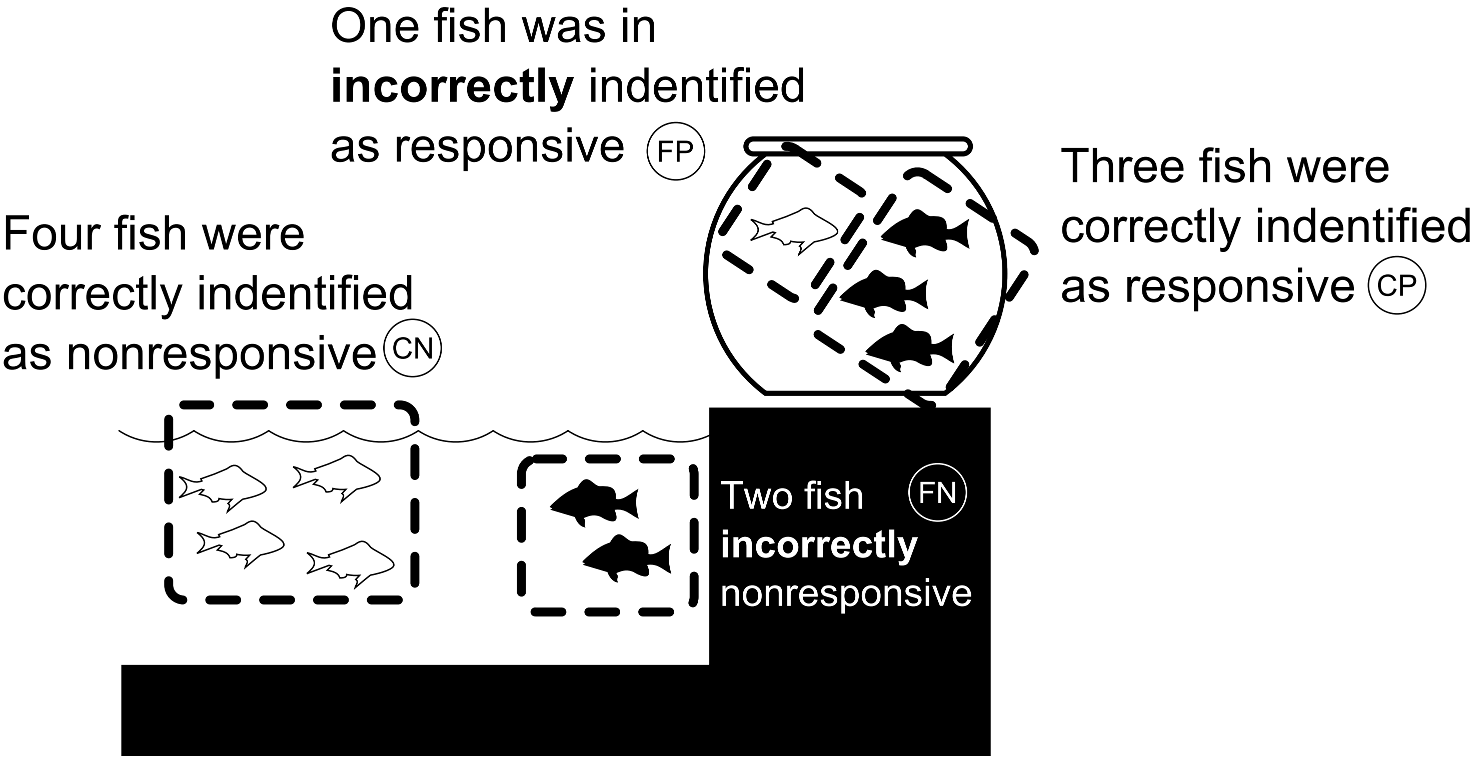
CP FP FN CN



Conclusion:

Reasonable minds
will disagree.

Understand what is happening here:



One fish was in **incorrectly** indentified as responsive (FP)

Three fish were correctly indentified as responsive (CP)

Four fish were correctly indentified as nonresponsive (CN)

Two fish (FN) **incorrectly** nonresponsive

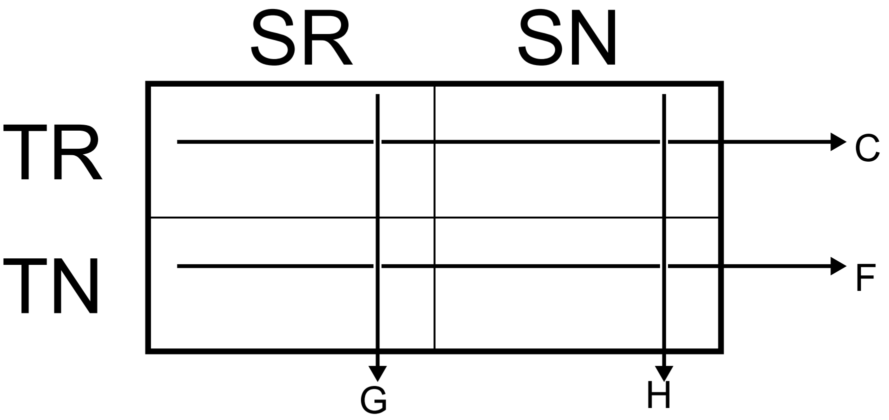
(A, B, D, and E.)

Overlay some codes to facilitate formulas (Don't be scared)

	SR	SN	
TR	(CP) 3 A	(FN) 2 B	5 C
TN	(FP) 1 D	(CN) 4 E	5 F
	4 G	6 H	10 I

	SR	SN	
TR	CP 3 A	FN 2 B	5 C
TN	FP 1 D	CN 4 E	5 F
	4 G	6 H	10 I

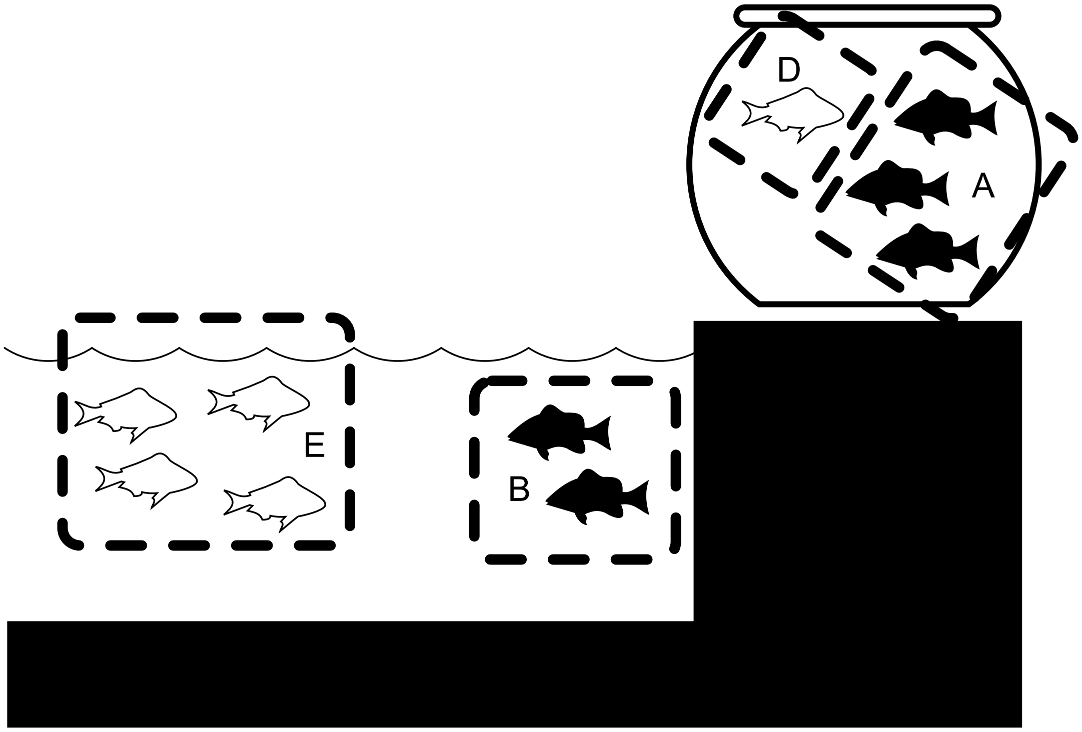
Realize that TR=C, TN=F, SR=G, and SN=H



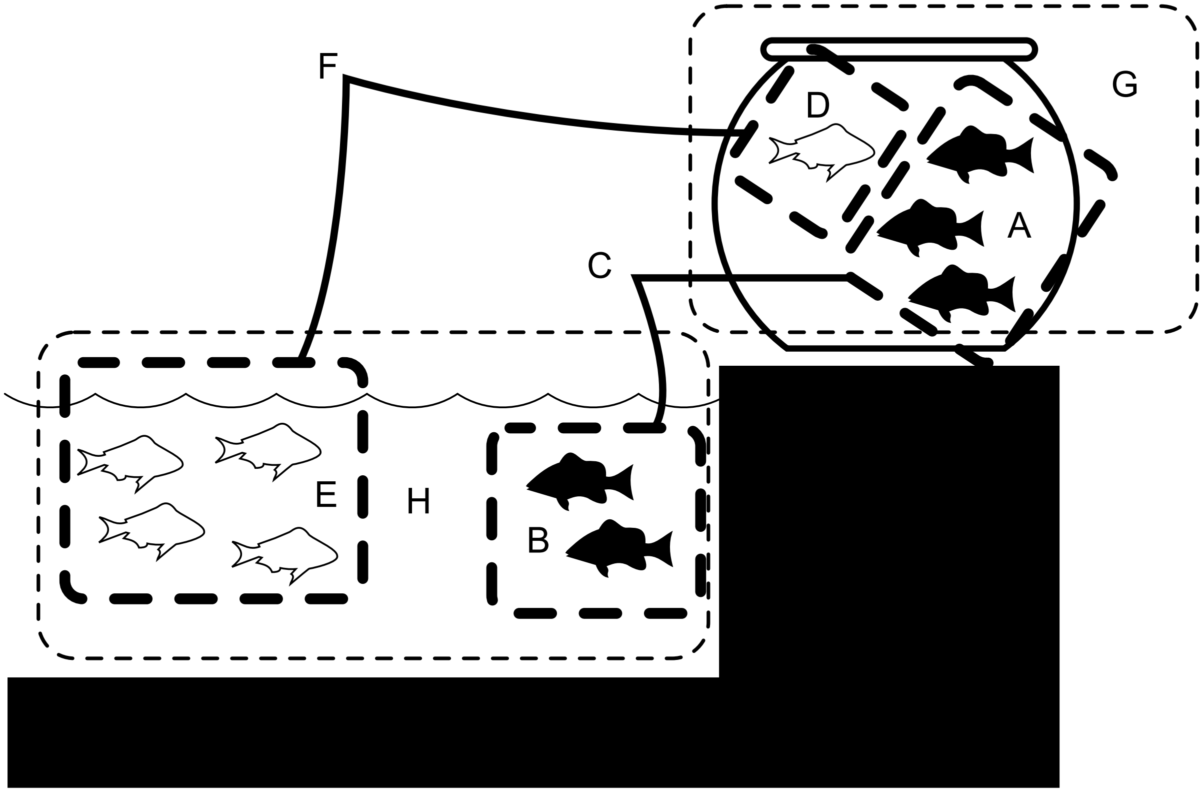
Each of nine values has a letter code, A-I

	SR	SN	
TR	A	B	C
TN	D	E	F
	G	H	I

Apply the interior codes (A, B, D, and E.) to the fish chart. Pretty simple.



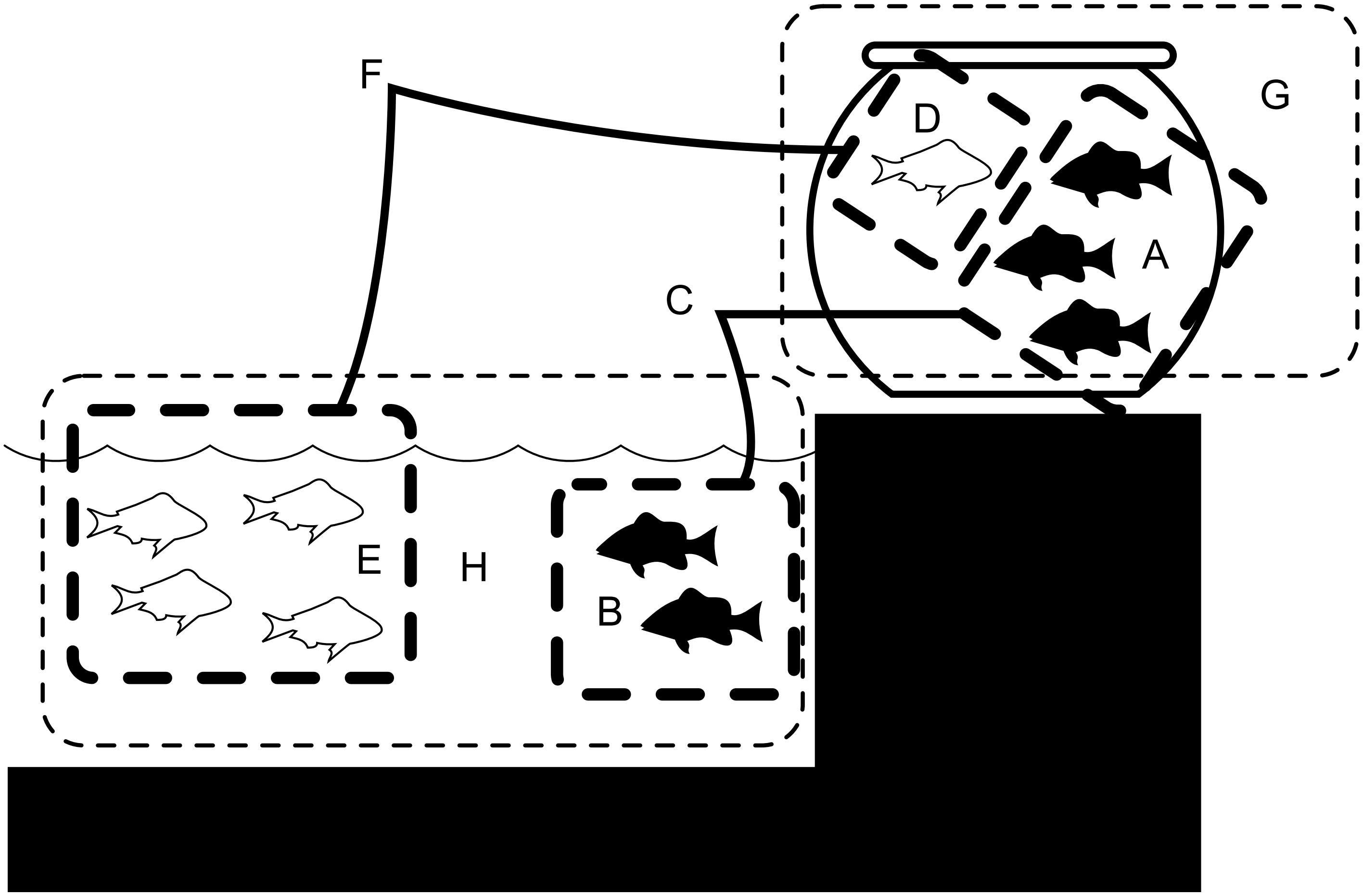
Now add the perimeter codes.



For the sake of exhausting repetition, here is the table and the diagram. Get used to them.

	SR	SN	
TR	3 A	2 B	5 C
TN	1 D	4 E	5 F
	4 G	6 H	10 I

Contingency Table filled out

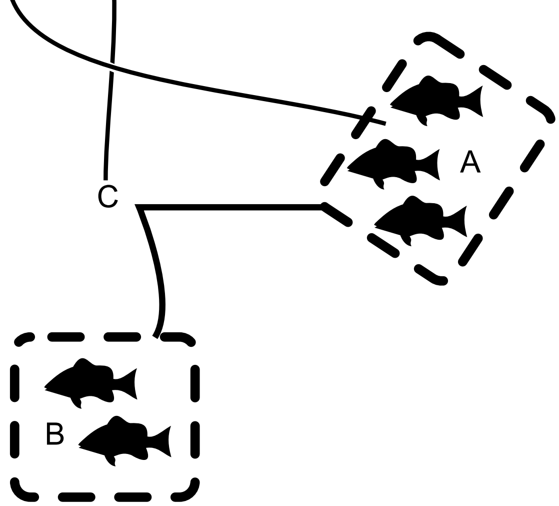


Now we will put together our first metric.

The most basic and argubly most important
for legal information seekers is RECALL.

RECALL is the percentage of responsive
documents that the search found.

Our document set has **five** responsive documents.
Our search found **three** of them.



RECALL = A/C

RECALL = 3/5 .6

Recall is the number of responsive documents in the search results divided by the total number of responsive documents in the complete document set

Recall is also know as:
True Positive Rate
Sensitivity
Hit Rate

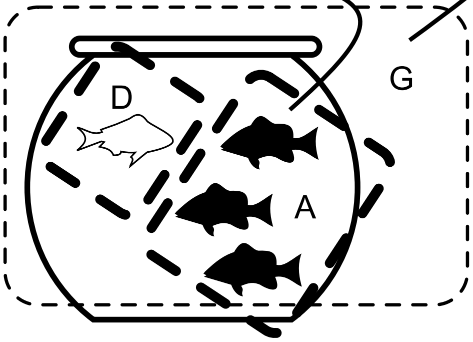
AKA Positive Predictive Value

Another basic and accessible formula is Precision. Precision is important for retrieval tasks such as internet searching.

PRECISION in the percentage of **retrieved** documents that are **responsive**.

Our search retrieved **four** documents. **Three** of them are responsive.

PRECISION = A/G



PRECISION = 3/4

.75

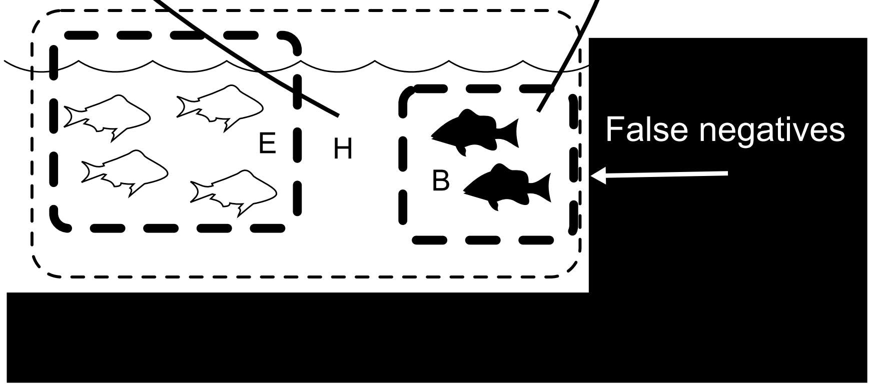
Precision is the number of responsive documents in the search results divided by the total number of documents in the search results.

ELUSION is the percentage of nonretrienved documents which are responsive and should have been retrieved.

Our document set has **six** nonretrieved documents. **Two** of them are responsive and should have been retrieved.

False negatives

ELUSION = B/H



ELUSION = 2/6 .33

Elusion allows us to assess whether our entire process has succeeded to the required level.

Baron, J. R & Thompson, P., Proceedings of the
11th international conference on
Artificial intelligence and law. 2007

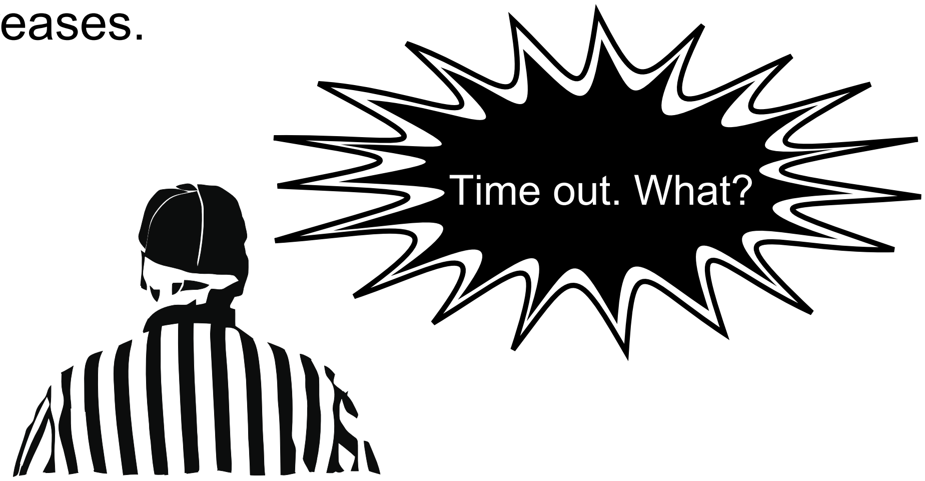
Proportion of predicted negatives that are incorrect.

Search nonresponsives that are responsive

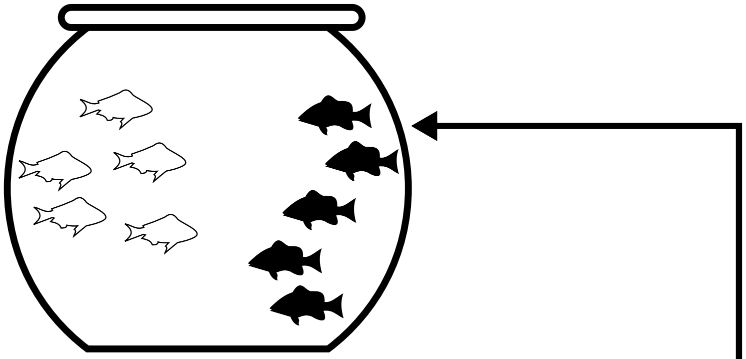
Instead of counting the responsive documents that we found, we count the ones that we left behind.

H. L. Roitblat, Measurement in eDiscovery
2013 OrcaTec LLC

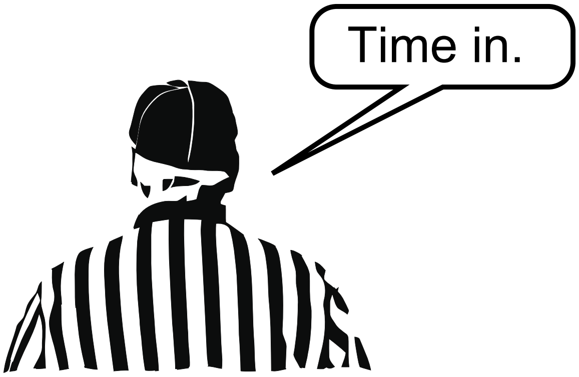
FALLOUT measures how quickly PRECISION drops as RECALL increases.



If I want to increase my recall, I need to get more black fish in the bowl. So I adjust my search. I get maximum recall by leaving no black fish behind. So maybe I set my search so that EVERY fish is caught.
100% RECALL. But what about PRECISION?



Remember that PRECISION in the percentage of retrieved documents that are responsive. So in this example, PRECISION dropped to 50%.



FALLOUT measures how quickly PRECISION drops as RECALL increases.

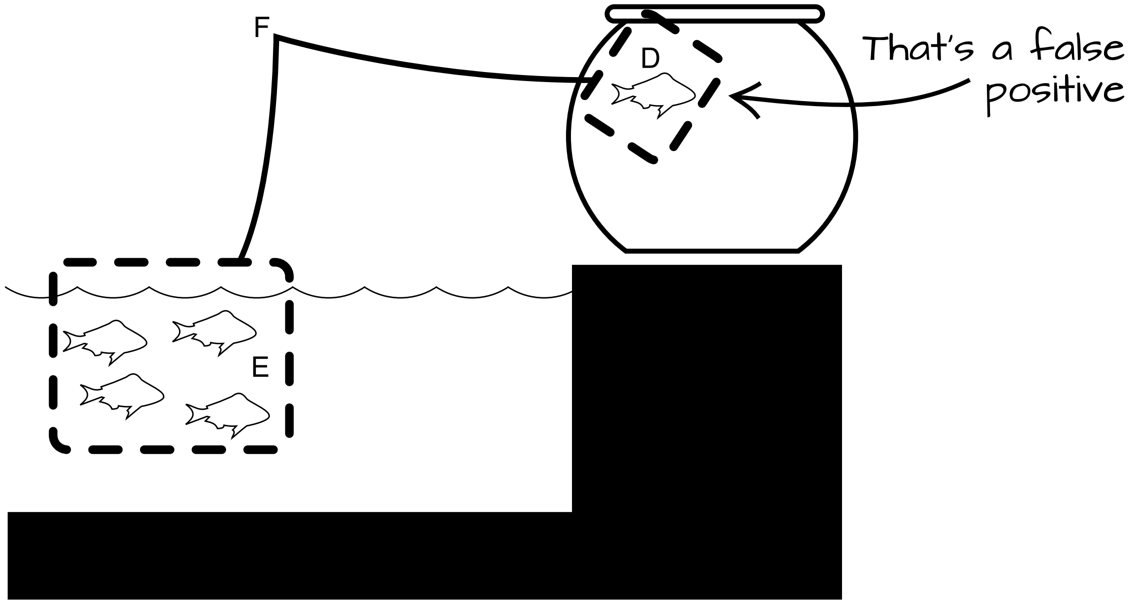
FALLOUT is the percentage of all nonresponsive documents which were incorrectly retrieved.

white fish

Our document set has **five** nonresponsive documents.
Our search incorrectly found **one** of them.

white fish in the bowl

FALLOUT = D/F



FALLOUT = 1/5 .2

Be careful with fallout because you can easily get a fallout of zero by marking zero documents responsive.

Before we look at another formula, let's look back at precision and recall.

1 **Start** with the full loaded contingency table

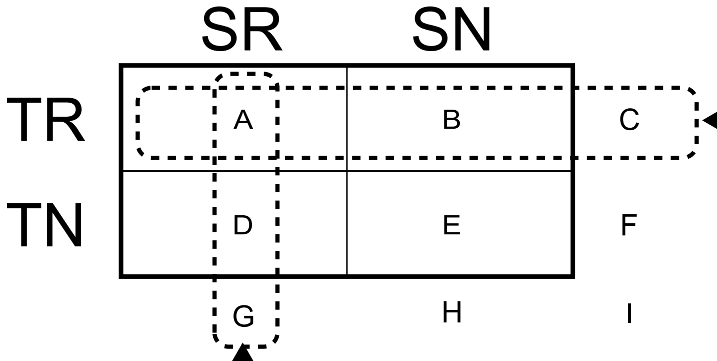
	SR	SN	
TR	CP 3 A	FN 2 B	5 C
TN	FP 1 D	CN 4 E	5 F
	4 G	6 H	10 I

2 **Remember** the recall and precision formulas:

RECALL = A/C

PRECISION = A/G

3 **Notice** the rows and columns that the formulas draw from:

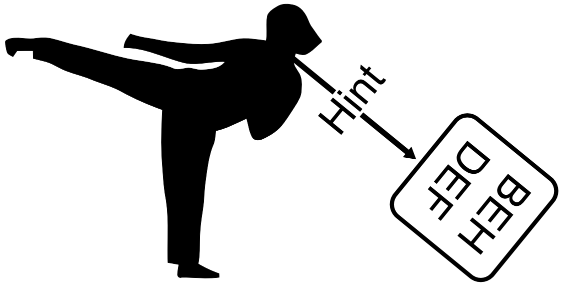


	SR	SN	
TR	A	B	C
TN	D	E	F
	G	H	I

4 **Do** the same thing for Elusion and Fallout.

If that was obvious to you and you did not need the diagram, this is for you:

(for kicks)



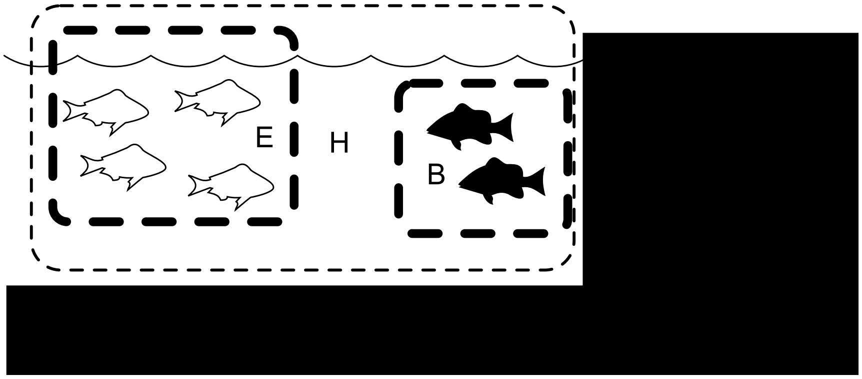
NEGATIVE PREDICTIVE VALUE reflects the percentage of non-retrieved documents that are in fact not responsive.

SN all fish in the water

Our search yielded **six** non-retrieved documents. Of these, **four** were not respsonsive.

CN white fish in the water

NPV = E/H

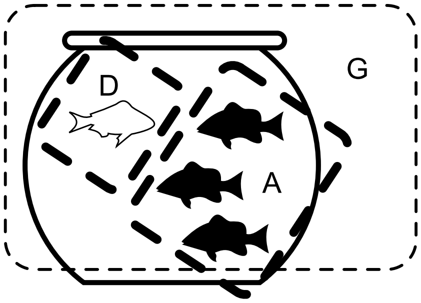


NPV = 4/6 .67

NPV is also 100% - ELUSION

minus

Note that NPV logically complements precision.



PRECISION = A/G

AKA yield AKA richness

PREVALENCE is the percentage of all documents which are true responsive.

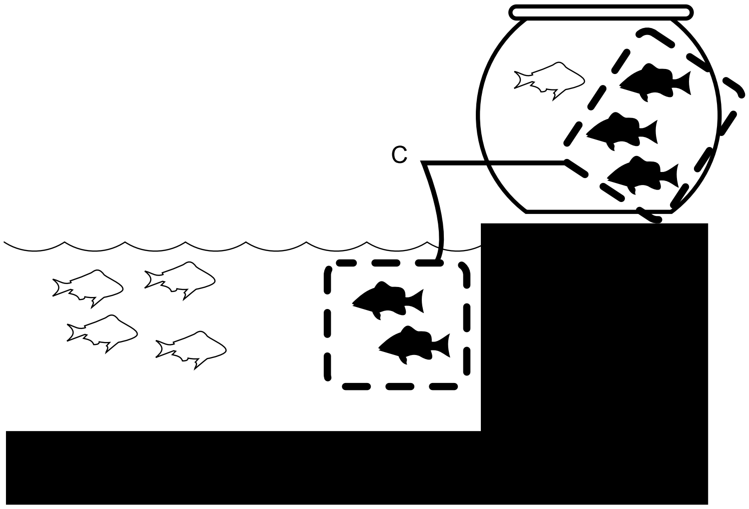
all fish

Our document set has **ten** documents.
Five are true respsonsive.

black fish

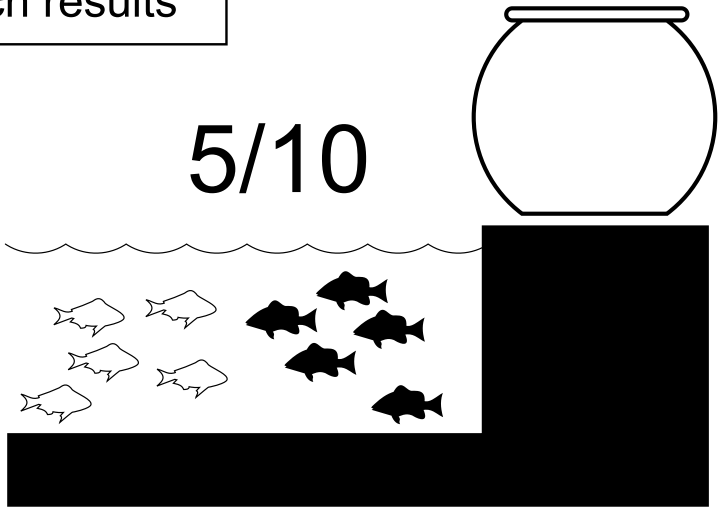
Prevalence = C\I

Prevalence = 5/10



NOTICE

This metric does not care about search results



SPECIFICITY is the percentage of true nonresponsive documents that are corrently identified as nonresponsive

white fish

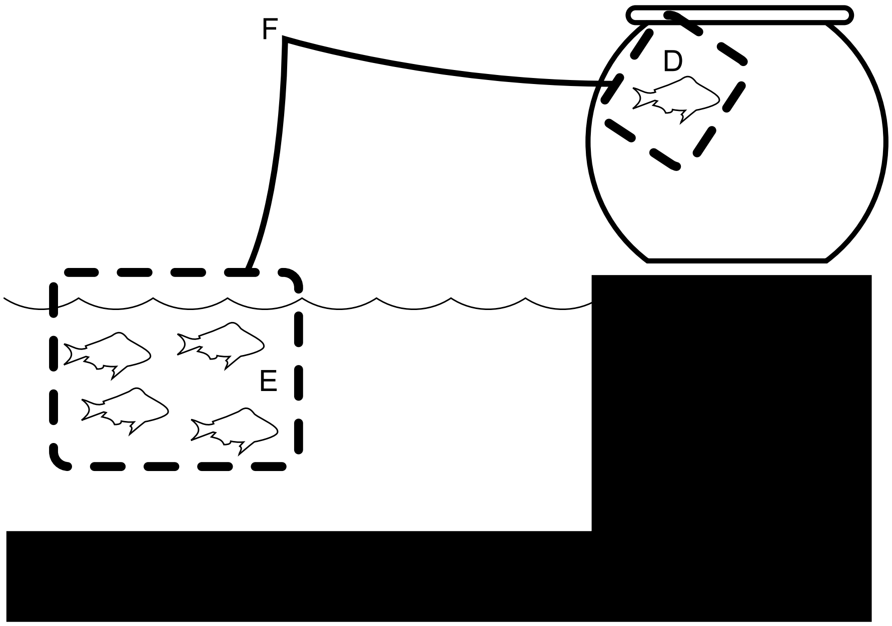
Our document set has **five** nonresponsive documents. **Four** were correctly identified.

white fish in the water

Specificity = E/F

AKA Correct Rejection Rate
AKA: True Negative Rate
AKA: Inverse Recall

Specificity = 4/5



compare this to fallout (same denominator, switch the numerator)

"the bottom number" "the top number"

credit: Black's Math Dictionary

FALSE NEGATIVE RATE

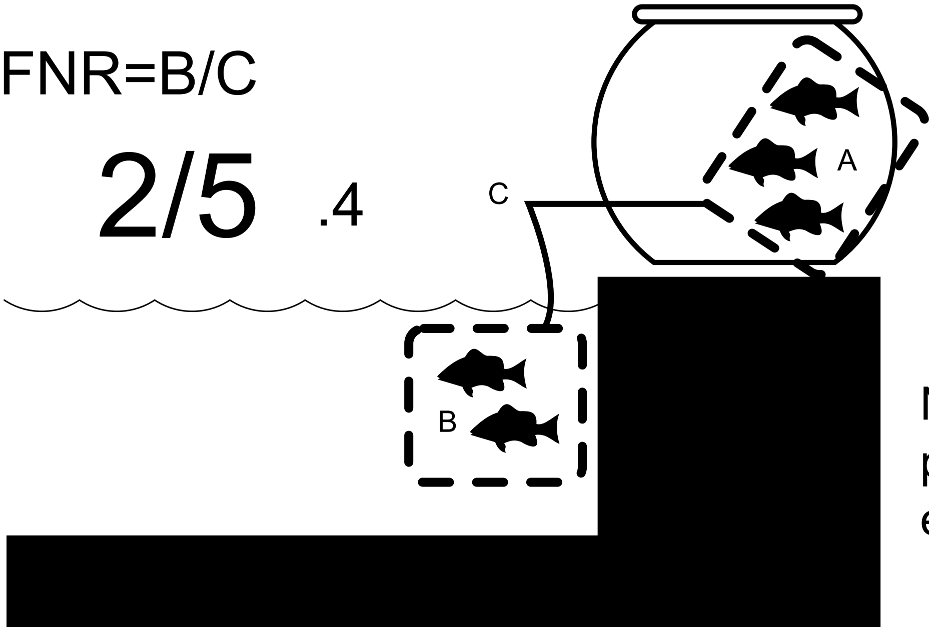
AKA Miss Rate

The percentage of True Responsive documents that are missed

100% – Recall True Positive Rate

FNR=B/C

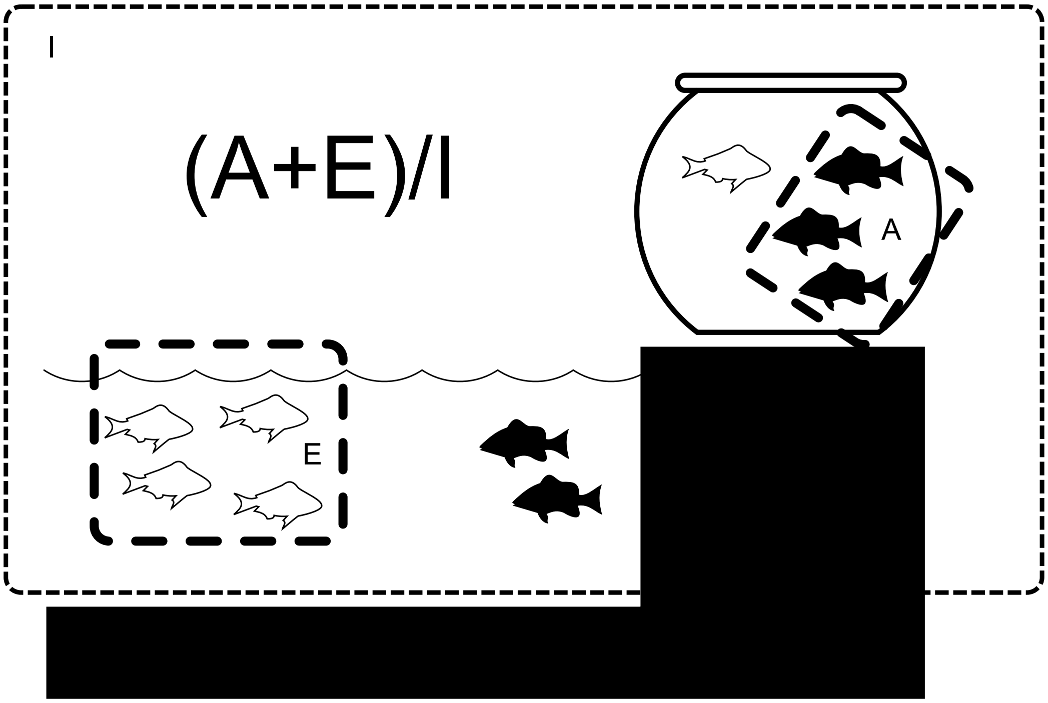
2/5 .4



Note that FNR plus Recall eqauls 100%

Accuracy

The percentage of documents that are correctly coded



$(A+E)/I$

$(3+4)/10$

70%

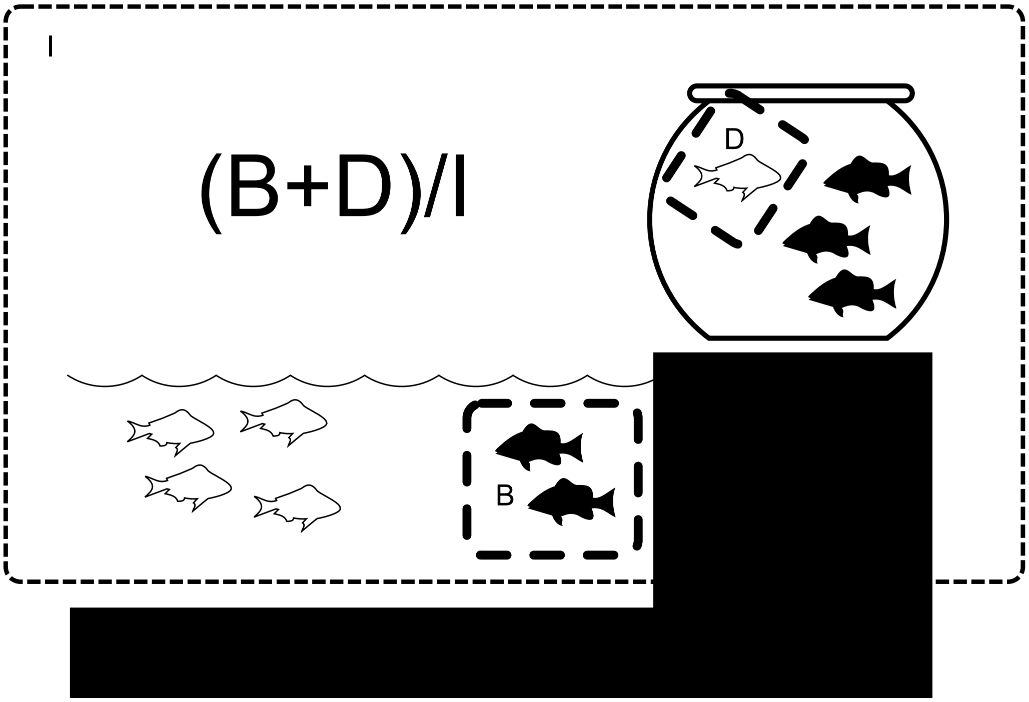
Accuracy is 100% - Error

minus

In highly prevalent or rich data sets (Or sets with extremely low prevalence or richness), Accuracy is a poor measure. Consider a set with 95 percent nonresponsive documents - 95 percent accuracy can be achieved by marking everything nonresponsive.

Error

The percentage of documents that are incorrectly coded



$(B+D)/I$

$(2+1)/10$

30%

Error can also be calculated: 100% – Accuracy

minus

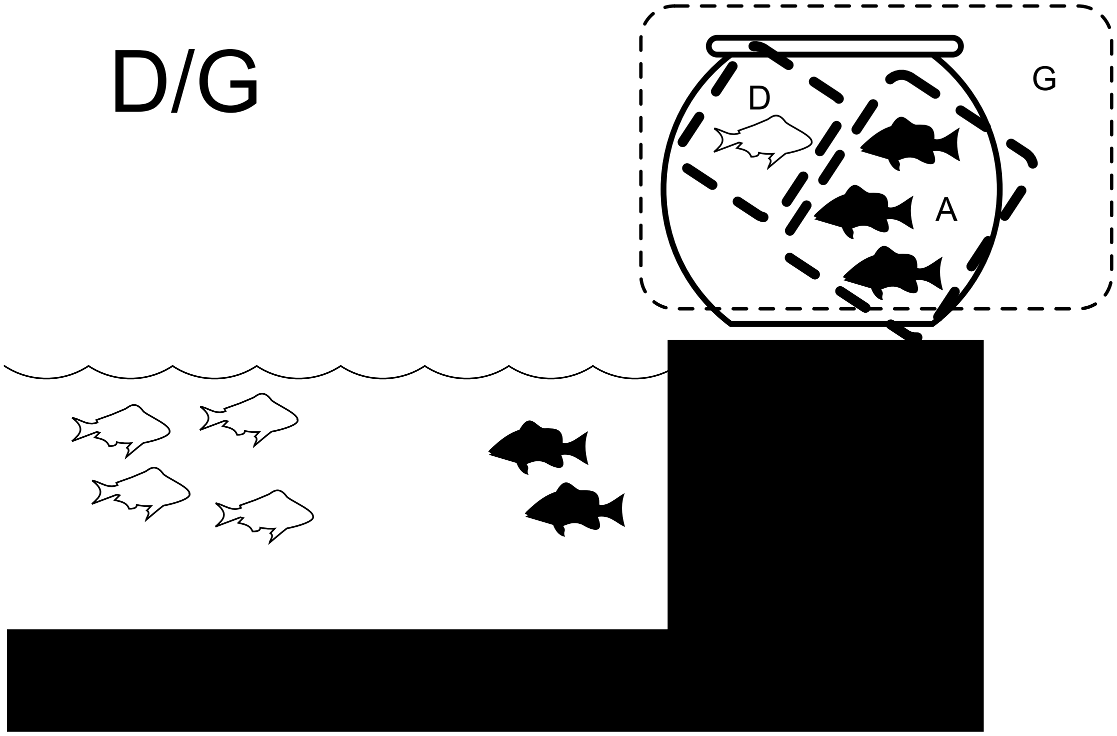
The warning regarding extremes of prevalence or richness applies to Error as well. The utility of Error as a search metric goes down as richness gets extremely high or low.

False Positive Rate

Flase Alarm Rate

The percentage of Search Responsive documents that are truly nonresponsive.

D/G



This metric does not care about the null set.