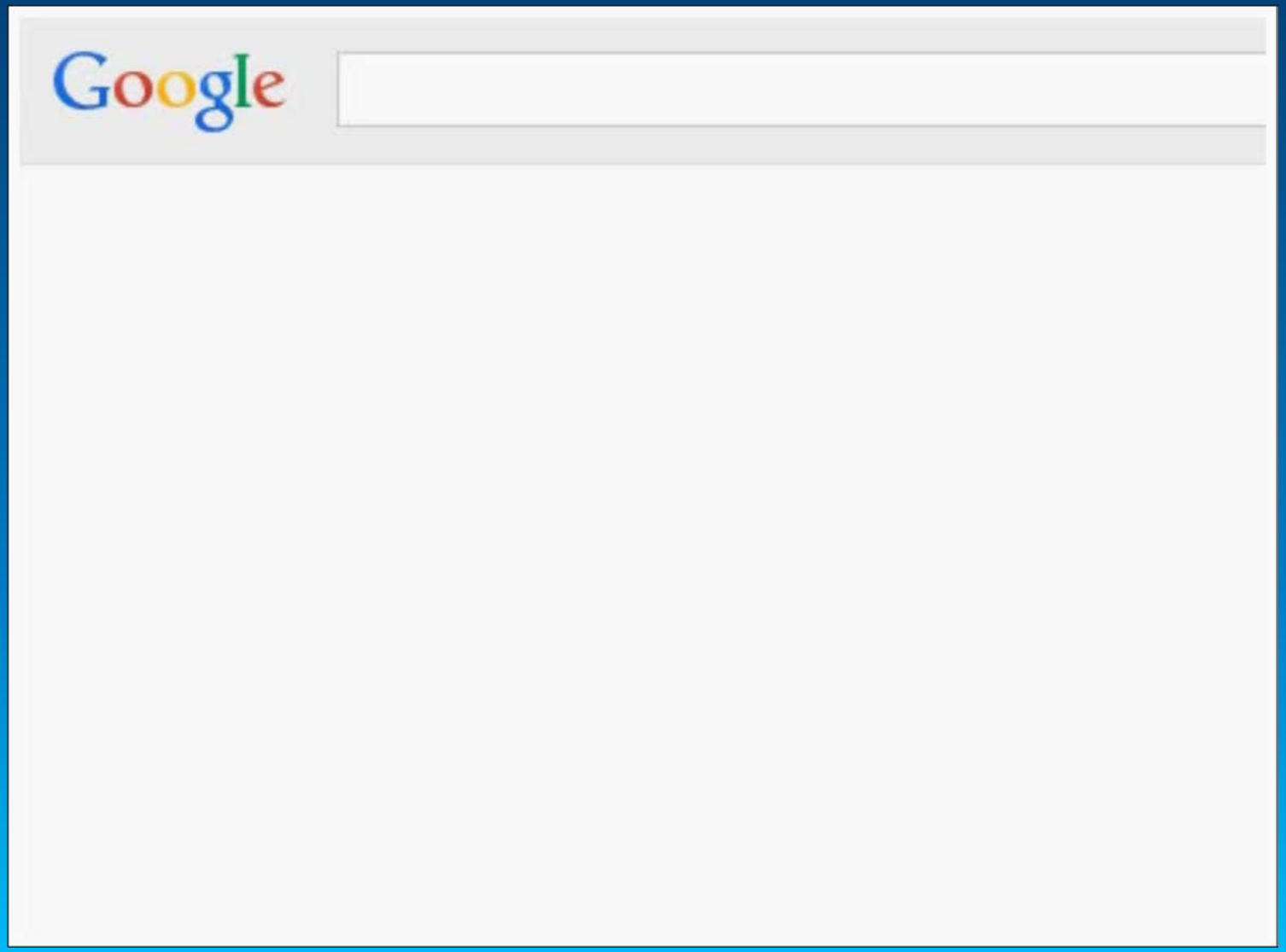
Create Fast Searches for Your Applications

David Haines, Boulder County Land Use

“Can you make it work like Google?”



Search Design Rules

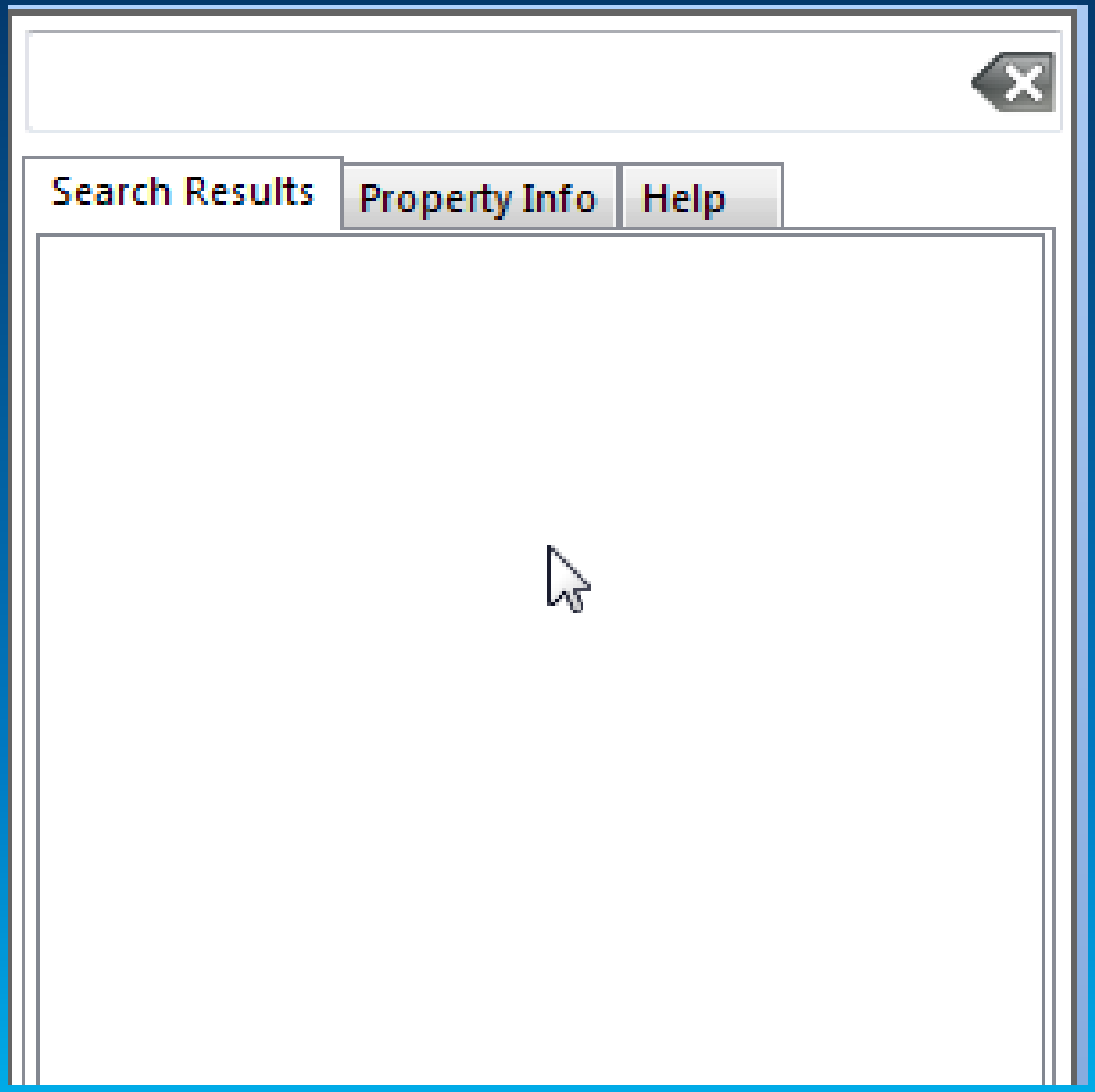
#1. Auto-complete

#2. One search field

#3. Without at least one of the above people will (or try to) use something else.

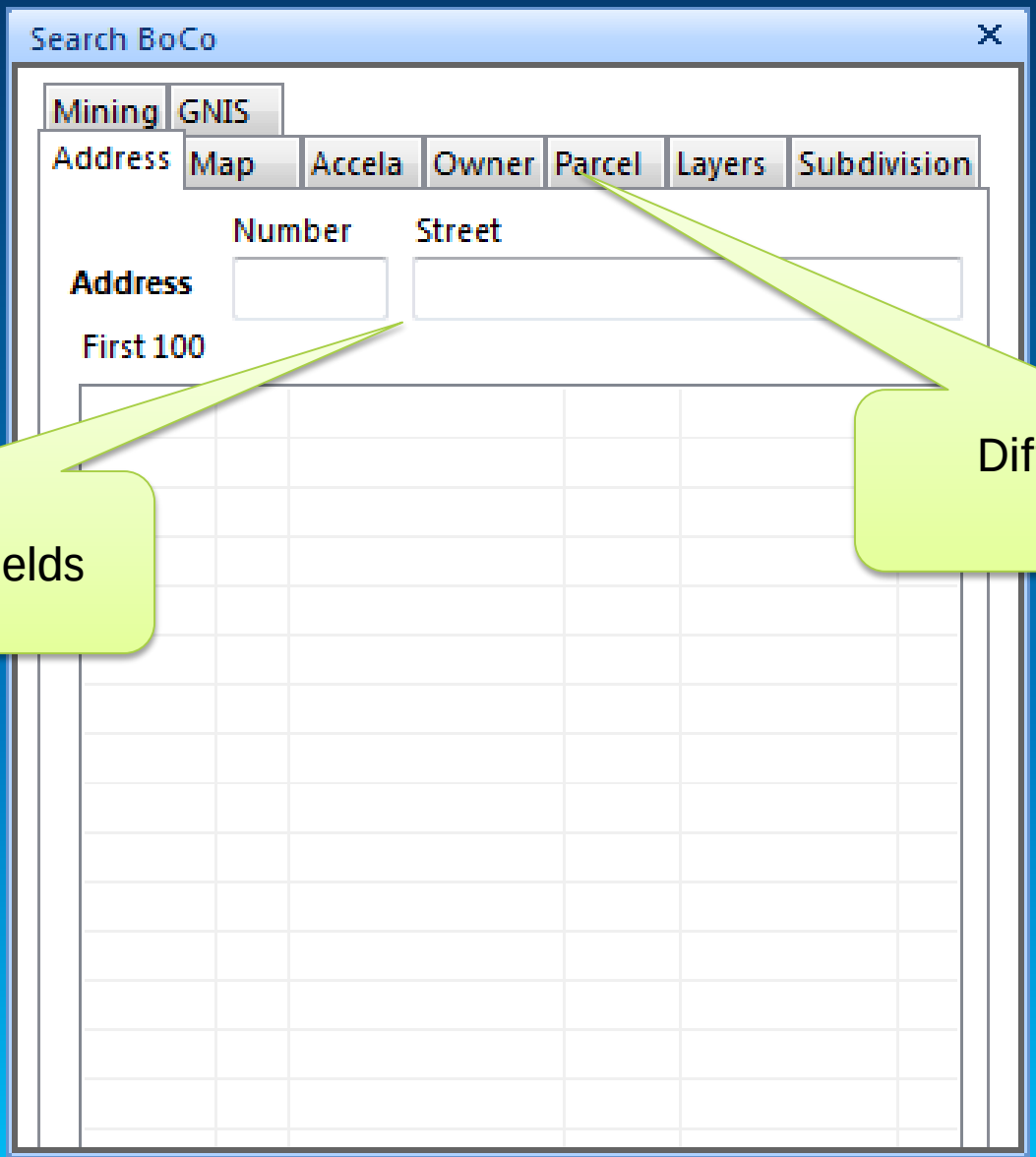
#4. Good results

Search Demo



Wasn’t Always this Way

Old Search Tool



Multiple Entry Fields

Different Types of Search

Too Many Databases

Hay for the Winter by Trey Ratcliff
http://www.flickr.com/photos/stuckincustoms/8672061349/
Attribution License

Address & Owner Database

Building Permit Database

Assessment Database

Place Name Database

Subdivision Database

Hay for the Winter by Trey Ratcliff
http://www.flickr.com/photos/stuckincustoms/8672061349/
Attribution License

One BIG Haystack

Roll of Hay by Klearchos Kapoutsis
http://www.flickr.com/photos/klearchos/3824322183/
Attribution-NonCommercial License

Roll of Hay by Klearchos Kapoutsis
http://www.flickr.com/photos/klearchos/3824322183/
Attribution-NonCommercial License

A Table To Search

What's in the Table?

What to search for

Where to go once you find it

Search Index Table Fields

Field	Description
SearchText	The search term. The “correct” answer. A perfect hit.

Search Index Table Fields

Field	Description
SearchText	The search term. The “correct” answer. A perfect hit.
Workspace	The database the search term is in

Search Index Table Fields

Field	Description
SearchText	The search term. The “correct” answer. A perfect hit.
Workspace	The database the search term is in
FeatureClass	The table in the database the search term is in

Search Index Table Fields

Field	Description
SearchText	The search term. The “correct” answer. A perfect hit.
Workspace	The database the search term is in
FeatureClass	The table in the database the search term is in
IdField	The field in the table the search term is in

Search Index Table Fields

Field	Description
SearchText	The search term. The “correct” answer. A perfect hit.
Workspace	The database the search term is in
FeatureClass	The table in the database the search term is in
IdField	The field in the table the search term is in
Id	The record (or row) the the search term is in. This is what the search should go to.

1234 Main St

1234 Main St

Match

Field	Description
SearchText	**1234 S Main Street Boulder**
Workspace	G:\gis.sde
FeatureClass	dbo.parcel
IdField	PIN
Id	MKE11171971

1234 Main St

Field	Description
SearchText	1234 S Main Street Boulder
Workspace	G:\gis.sde
FeatureClass	dbo.parcel
IdField	PIN
Id	**MKE11171971**

Steps to Make the Index Table

Boulder County Land Use Search Index

Theme	Example	Database
Parcel Number	157416001234	Assessor - CAMA
Address	1234 S Main Street	Assessor – CAMA
Owner	John Doe	Assessor – CAMA
Tax Account	R01234567	Assessor – CAMA
Building Permit	BP-13-0001	Land Use - Accela
Docket Number	SPR-13-0010	Land Use – Accela
Docket Name	Smith Residence	Land Use – Accela
Subdivision	Big Oak Meadows	GIS - ArcSDE
Mining Claim Name	Blue Bird Mine #2	GIS – ArcSDE
Geographic Names	Longs Peak	GIS - ArcSDE

Steps to update the index table

Load existing index

Steps to update the index table

Load existing index

Load reference data (i.e. data your searching)

Steps to update the index table

Load existing index

Load reference data (i.e. data your searching)

Find reference data not in index

Steps to update the index table

Load existing index

Load reference data (i.e. data your searching)

Find reference data not in index

Add that data

Steps to update the index table

Load existing index

Load reference data (i.e. data your searching)

Find reference data not in index

Add that data

Search for index data not in the index (deleted)

Steps to update the index table

Load existing index

Load reference data (i.e. data your searching)

Find reference data not in index

Add that data

Search for index data not in the index (deleted)

Remove it

Steps to update the index table

Load existing index

Load reference data (i.e. data your searching)

Find reference data not in index

Add that data

Search for index data not in the index (deleted)

Remove it

Repeat for each dataset

One Search Index

Many Applications

Gerber Multi-Plier, model 400 Compact Sport, shown with tools open by Fluzwup
https://en.wikipedia.org/wiki/File:Mp_400_sport.jpg
CC **Attribution-ShareAlike**

Other Tips

(If Time Allows)

Optimizing wildcard searching

Word1 Word2%

Word1% %Word2%

Word1% %Word2% %Word3% …

%Word1% %Word2%

%Word1% %Word2% %Word3% …

Score Your Results

The Levenshtein distance between two words is the minimum number of single-character edits (insertion, deletion, substitution) [including spaces] required to change one word into the other.

1234 Ced

1234 N Cedar Brook Lane

Source: en.wikipedia.org/wiki/Levenshtein_distance

Trigrams!

THE QUICK RED FOX

THE

HE_

E_Q

_QU

QUI

UIC

ICK

CK_

K_R

_RE

Regular Expression Matching with a Trigram Index or How Google Code Search Worked
Russ Cox
https://swtch.com/~rsc/regexp/regexp4.html

Questions?

dhaines@bouldercounty.org