

From: Hoy, Mark
To: [Bowman, Randal](#)
Cc: stu@texifter.com; [Cash, Marcia](#)
Subject: Re: Status on current upload?
Date: Thursday, July 27, 2017 1:26:18 PM
Attachments: [indexStopwords.txt](#)

Randal -

I'm not entirely sure that the stopwords are indeed the problem - I'm still researching this further. Some search engines will ignore stopwords in quoted (phrase) queries, and some will not... I've not yet been able to find a definitive answer for how our search engine (3rd party) treats stop words within quoted queries... my best guess so far is that _yes_ this is indeed what is going on, but I'm still not 100% convinced as I cannot find any references saying "yes or no" to if they are ignored or not.

For your reference, the list of stopwords in use by our search engine is attached.

- Mark

----- Forwarded message -----

From: Bowman, Randal <randal_bowman@ios.doi.gov>
Date: Thu, Jul 27, 2017 at 1:29 PM
Subject: Re: Status on current upload?
To: Stu Shulman <stu@texifter.com>
CC: Marcia Cash <marcia_cash@fws.gov>

My concern was and is that I did a search and got results not corresponding to the search term, which caused me to stop searches and stop coding the results.

This is the first I've heard of "stopwords". If the problem is that some words, even when used with other words and in quotes, don't register with discovertext searches, a simple explanation of that with a list of the words or types of words would have resolved the matter.

On Thu, Jul 27, 2017 at 12:54 PM, Stu Shulman <stu@texifter.com> wrote:
From Mark:

I don't know why he's still asking about the quotes... I thought I made it clear that the system has been patched (for a couple of weeks) and will transform the curly quotes into regular quotes... :/
As for his exact issue (I _think_ he's saying that he is seeing search results when searching for "mounument as is" that doesn't have the phrase "as is" in it... .) - I _think_ this is related to "as" and "is" being stopwords (and thus, not counted in the query), but I've not been able to determine this 100% yet...

On Thu, Jul 27, 2017 at 12:26 PM Bowman, Randal <randal_bowman@ios.doi.gov> wrote:

How is this coming? I will shortly send a note to the DOI coders urging them to finish the current project, where there are about 3200 comments remaining to be coded, so we can focus on the new ones next week, and would like to include when we likely can start that.

Also, I know you addressed this early on, but I am not sure I remember it correctly - can I only access my account from my government computer, or will the link with my username and password work from any computer? If the latter, I likely would have some time available while at my parents'.

Lastly, has Mark had any chance to work on the quotes/search issue, or has all his time been taken on getting the new comments loaded?

--

Dr. Stuart W. Shulman
Founder and CEO, Texifter
Cell: [413-992-8513](tel:413-992-8513)
LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

--

Dr. Stuart W. Shulman
Founder and CEO, Texifter
Cell: [413-992-8513](tel:413-992-8513)
LinkedIn: <http://www.linkedin.com/in/stuartwshulman>

0
1
2
3
4
5
6
7
8
9
a
about
above
across
after
again
against
al
all
almost
alone
along
already
also
although
always
am
among
an
and
another
any
anybody
anyone
anything
anywhere
are
area
areas
around
as
ask
asked
asking
asks
at
away
b
back
backed
backing
backs
be
became
because
become
becomes
been
before
began
behind
being

beings
best
better
between
big
bitly
both
but
by
c
came
can
cannot
case
cases
certain
certainly
clear
clearly
co
com
come
como
con
could
d
da
de
del
did
differ
different
differently
do
does
done
down
down
downed
downing
downs
during
e
each
early
edu
either
el
en
end
ended
ending
ends
enough
es
even
evenly
ever
every
everybody
everyone
everything

everywhere
f
face
faces
fact
facts
far
felt
few
find
finds
first
for
four
from
full
fully
further
furthered
furthering
furthers
g
gave
general
generally
get
gets
give
given
gives
go
going
good
goods
got
gran
great
greater
greatest
group
grouped
grouping
groups
h
had
has
have
having
he
her
here
herself
high
high
high
higher
highest
him
himself
his
how
however

http
https
i
if
im
important
in
interest
interested
interesting
interests
into
is
it
its
itself
ive
j
just
k
keep
keeps
kind
knew
know
known
knows
l
la
large
largely
las
last
later
latest
least
less
let
lets
like
likely
lo
long
longer
longest
los
m
made
make
making
man
many
may
me
member
members
men
might
more
most
mostly
mr

mrs
much
must
my
myself
n
necessary
need
needed
needing
needs
never
new
new
newer
newest
next
no
nobody
non
noone
not
nothing
now
nowhere
number
numbers
o
of
off
often
oh
old
older
oldest
on
once
one
only
open
opened
opening
opens
or
order
ordered
ordering
orders
org
other
others
our
out
over
p
para
part
parted
parting
parts
per
perhaps

place
places
point
pointed
pointing
points
por
possible
present
presented
presenting
presents
problem
problems
put
puts
q
que
quite
r
rather
really
right
right
room
rooms
rt
s
said
same
saw
say
says
se
second
seconds
see
seem
seemed
seeming
seems
sees
seu
several
si
shall
she
should
show
showed
showing
shows
side
sides
since
small
smaller
smallest
so
some
somebody
someone

something
somewhere
state
states
still
still
su
such
sure
t
take
taken
than
that
the
their
them
then
there
therefore
these
they
thing
things
think
thinks
this
those
though
thought
thoughts
three
through
thus
to
today
together
too
took
toward
turn
turned
turning
turns
two
u
un
una
under
until
up
upon
us
use
used
uses
v
ver
very
w
want
wanted

wanting
wants
was
way
ways
we
well
wells
went
were
what
when
where
whether
which
while
who
whole
whose
why
will
with
within
without
work
worked
working
works
would
x
y
ya
year
years
yet
you
young
younger
youngest
your
yours
z