

Re: display hit results in proper format

Source:

<http://www.tech-archive.net/Archive/Internet-Server/microsoft.public.inetserver.indexserver/2006-10/msg00029.htm>

- *From:* "Hilary Cotter" <hilary.cotter@xxxxxxxx>
 - *Date:* Thu, 26 Oct 2006 11:40:32 -0400
-

1) its done at query time. The text is extracted from the file and marked up and displayed on the web. IIRC They launch the ifilter for the document type in process and use that to extract the text.

2) The problem is that you really need to use the Microsoft word breaker/stemmers to correctly mark up freetext searches on mice to match with mouse. A grep or regex won't be able to do this, but the Microsoft stemmers/word breakers do – especially for non English languages where the endings get even more irregular.

—
Hilary Cotter
Director of Text Mining and Database Strategy
RelevantNOISE.Com – Dedicated to mining blogs for business intelligence.

This posting is my own and doesn't necessarily represent RelevantNoise's positions, strategies or opinions.

Looking for a SQL Server replication book?
<http://www.nwsu.com/0974973602.html>

Looking for a FAQ on Indexing Services/SQL FTS
<http://www.indexserverfaq.com>

"Steve" <dudesterr@xxxxxxxx> wrote in message
<news:1161867760.944861.49500@xx>

Hey Hilary
thanks for the reply,
hey you're amazing,i've been searching through the posts and wow i'm really head over heels in here ,you're everywhere !!
save the boys some,will ya !!
ok now back to business
i got some inquiries about this issue

1– format of the result returned by webhits.dll is just simple html

Re: display hit results in proper format

right but how is it converted from all kind of files into html or is that done at indexing time ?

2- i wrote a hit highlighter script that works only on html files, and i intend to write other types of highlighters as well ie. for pdf or other .. so i might need some guidance in here in terms of where to look and what to look for and if you know some ready product i'll be very grateful.

thanks
keep up the good work