Tech Info Library

AppleSearch 1.5 and Searches with Wildcards (11/95)

Revised: 1/4/96
Security: Everyone

AppleSearch 1.5 and Searches with Wildcards (11/95)

==

Article Created: 24 March 1995
Article Rewviewed/Updated: 20 November 1995

TOPIC ---

We have an issue with AppleSearch using wildcards. We are attempting to put our
HR records into AppleSearch. We have noticed a number of occurrences of
wildcards not working very well. One glaring example is the variations of the
word manage. For example, management will not be found with any wildcards except
for managemen*, and even then does not find many of the hits that using
management will find. What am I doing wrong?

DISCUSSION --

At the heart of this issue is the stemming technique used by AppleSearch for
search strings. When you enter a wildcard string, you expect AppleSearch to use
the entire string, as entered, and the wildcard to find matches. Unfortunately
AppleSearch does not do this. It first stems the string and then places the
wildcard. It then uses this string to search the index, where the index is
composed of stemmed references to document text. This is how a search for
"representing" can find such words like represent, represents, and
representative. AppleSearch stems representing to "repres", as do the words
represent, represents, and representative. In essence, AppleSearch believes
these words are the same, because its index only contains the stemming results,
not the whole string.

Unfortunately, stemming is not always a predictable science. From your example,
we will use the manage* string. Stemming of this string produces "manage*".
However, many of the words desired by the search (such as management, manager,
managing, and so on) stem to manag, which is why many of the expected hits are
not returned.

Here is a general guideline that should be useful for users that are having
problems with search strings:

In general, wildcards are not necessary for AppleSearch queries as stemming

provides the same results. When wildcards are unavoidable, try to make the provided information as small as possible, usually four or five characters maximum.

Article Change History:
20 Nov 1995 - added new keyword and reviewed for technical accuracy.
Support Information Services

Tech Info Library Article Number:17424