DICOM Correction Item

Correction Number	CP-899

| Log Summary: Matching of multi-component names ||

Type of Modification	Name of Standard
Clarification	PS 3.4 2008

Rationale for Correction

There is no special behavior defined for matching of multi-component names such as those with ideographic and phonetic components. Clarify that literal string matching may give unexpected results, both for single value and wildcard matching.

Sections of documents affected

PS 3.4

Correction Wording:

Amend PS 3.4 C.2.2.2 Attribute Matching:

C.2.2.2   Attribute Matching

The following types of matching may be performed on Key Attributes in the Query/Retrieve Service Class:

— Single Value Matching
— List of UID Matching
— Universal Matching
— Wild Card Matching
— Range Matching
— Sequence Matching

Matching requires special characters (i.e. "*", "?", "-", **"="** and "\") which need not be part of the character repertoire for the VR of the Key Attributes.

> Notes: 1.For example, the "-" character is not valid for the DA, DT and TM Vrs but is used for range matching.
>
> 2. When character sets other than the default character repertoire are used, then the rules in PS 3.5 apply, such as with respect to the use of the 05/12 "\" (BACKSLASH) (in ISO IR 6) or 05/12 "¥" (YEN SIGN) (in ISO IR 14).

The total length of the Key Attribute may exceed the length as specified in the VR in PS 3.5. The Value Multiplicity (VM) may be larger than the VM specified in PS 3.6 for the Key Attribute, as defined for particular Matching Type.

The Specific Character Set (0008,0005) Attribute may be present in the Identifier but is never matched. Rather, it specifies how other Attributes are encoded in the Request and Response Identifiers.

It may influence how matching of other Attributes is performed. If Specific Character Set (0008,0005) is absent, then the default character repertoire shall be used. Specific Character Set (0008,0005) shall not have a zero length value.

Specific Character Set (0008,0005) may have multiple values if escape sequences are used to switch between character repertoires within values.

If the SCP does not support the value(s) of Specific Character Set (0008,0005) in the Request Identifier, then the manner in which matching is performed is undefined and shall be specified in the conformance statement.

Note**s**:   **1.** If an SCU sends a Request Identifier with a single byte character set not supported by the SCP, then it is likely, but not required, that the SCP will treat unrecognized characters as wildcards and match only on characters in the default repertoire, and return a response in the default repertoire.

2. Some Specific Character Set values are used with multi-component group person names (e.g., single-byte, ideographic and phonetic and phonetic component groups separated by an "=" (3DH) character), which may also affect the behavior of literal string matching.

C.2.2.2.1   Single Value Matching

If the value specified for a Key Attribute in a request is non-zero length and if it is:

a)   not a date or time or datetime, contains no wild card characters
b)   a date or time or datetime, contains a single date or time or datetime with no "-"

then single value matching shall be performed. Except for Attributes with a PN Value Representation, only entities with values which match exactly the value specified in the request shall match. This matching is case-sensitive, i.e., sensitive to the exact encoding of the key attribute value in character sets where a letter may have multiple encodings (e.g., based on its case, its position in a word, or whether it is accented).

For Attributes with a PN Value Representation (e.g., Patient Name (0010,0010)), an application may perform literal matching that is either case-sensitive, or that is insensitive to some or all aspects of case, position, accent, or other character encoding variants.

Note: **For multi-component names, the component group delimiter "=" (3DH) may be present in the Key Attribute value, but may give unexpected results if the SCP does not support matching on separate components but interprets the entire value literally as a single string. E.g., "Wang^XiaoDong=王^小東" may or may not match "Wang^XiaoDong" or "王^小東"; wildcard matching without the component group delimiter, such as "*Wang^XiaoDong*" or "*王^小東*" may be necessary.**

If extended negotiation of fuzzy semantic matching rather than literal matching of PN Value Representation is successful, not only may matching be insensitive to case, position, accent, and character encoding, but in addition other techniques such as phonetic matching may be applied.

The manner in which matching is performed is implementation dependent and shall be specified in the conformance statement.

Notes: 1. This definition implies that dates or times or datetimes are matched by their meaning, not as literal strings. For example:
— the DT "19980128103000.0000" matches "19980128103000"
— the DT "19980128103000" matches "19980128073000-0300"
— the TM "2230" matches "223000"
— the TM "223000" matches the deprecated ACR/NEMA 2.0 form "22:30:00"
— the DA "19980128" matches the deprecated ACR/NEMA 2.0 form "1998.01.28"

2. If an application is concerned about how single value matching of dates and times is performed by another application, it may consider using range matching instead, which is always performed by meaning, with both values in the range the same.

3. Exclusion of the "-" character for single value matching implies that a Key Attribute with DT Value Representation may not contain a negative offset from Universal Coordinated Time (UTC) if single value matching is intended. Use of the "-" character in date, time or datetime indicates range matching.

4. If an application is in a local time zone that has a negative offset then it cannot perform single value matching using a local time notation. Instead, it can convert the Key Attribute value to UTC and use an explicit suffix of "+0000".

5. Matching of PN Attributes may be accent-insensitive, as specified in the conformance statement. Accent-insensitive matching would successfully match, for instance, a query character "SMALL LETTER a" (06/01 in the default ISO-IR 6) with

"SMALL LETTER a WITH GRAVE ACCENT" (14/00 in ISO-IR 100),

"SMALL LETTER a WITH TILDE" (14/03 in ISO-IR 100),

"SMALL LETTER a WITH BREVE" (14/03 in ISO-IR 101), and

"CAPITAL LETTER a WITH ACUTE ACCENT" (12/01 in ISO-IR 100) (if matching is also case-insensitive),

but would not match 14/00 in ISO-IR 101, which is "SMALL LETTER r WITH ACUTE ACCENT". Matching to particular bit-combinations is specific to each supported character set (note the difference in meaning of 14/00), and should be described in the conformance statement.

6. An SCU application may elect to perform additional filtering of the responses by applying the matching rules itself. In the event that both the SCU and SCP are applying the matching rules, this process will be successful as long as literal matching is performed by both, and any additional SCU filtering is insensitive to case, position, accent, or other character encoding variants.

However if fuzzy semantic matching of PN Attributes has been negotiated, matching by the SCP may result in responses that are not obviously related to the request, hence care should be taken if any additional filtering of responses is performed by the SCU. For example, if phonetic matching is performed, a query for "Swain" might well return "Swayne", or if name component order insensitive matching is performed, a query for "Smith^Mary" might well return "Mary^Smith" or "Mary Smith" or "Smith, Mary". **Fuzzy semantic matching may also take into account separate single-byte, ideographic and phometic name component groups.**

...

C.2.2.2.4   Wild Card Matching

If the Attribute is not a date, time, signed long, signed short, unsigned short, unsigned long, floating point single, floating point double, other byte string, other word string, unknown, attribute tag, decimal string, integer string, age string or UID and the value specified in the request contains any occurrence of an "*" or a "?", then "*" shall match any sequence of characters (including a zero length value) and "?" shall match any single character. This matching is case sensitive, except for Attributes with an PN Value Representation (e.g., Patient Name (0010,0010)).

For attributes with a PN value representation, including the case of extended negotiation of fuzzy semantic matching, wild card matching is implementation dependent and shall be specified in the conformance statement.

Notes: 1. Wild card matching on a value of "*" is equivalent to universal matching.

2. The wild card matching method specified by DICOM might not be supported by some non-DICOM multi-byte character text processors.

3. For multi-component group names, of the component group delimiter "=" (3DH) may be present in the Key Attribute value, but may give unexpected results if the SCP does not support matching on separate components but interprets the entire value literally.

E.g., "*=*" or "*=*=*" may or may not return all strings, and hence is not equivalent to "*", nor to universal matching.