

Queries, Korean Style

Being a student of the Korean language and knowing of Unicode character sets for encoding non-English text, I wondered if Korean words (한국 말씀) could be stored and queried in an SQL database. I expected they could since Koreans use computer systems designed in the US and query their databases all the time. So I learned about Unicode and found that, on a Microsoft system anyway, you can create columns, tables, and whole databases with Korean names, spelled entirely in Korean characters. So I created the following table, 한국시 (Korean Poems), in SQL Server. This also gave me some practice with the IME input system:

가
내 고향 남쪽바다 그 파란 물 눈에 보이네; 꿈에 들 잊으리오 그 잔잔한 고향바다;
죽는 날까지 하늘을 우러러; 한점 부끄럼이 없기를; 잎새에 이는 바람에도; 나는 괴로워했다; 별을 노래하는 마음으로; 모든 죽어가는 것을 사랑해야지; 그리고 나한테 주어진 길을; 걸어가야겠다; 오늘밤에도 별이 바람에 스치운다;
나는 나룻배; 당신은 흙밭로 나를 깃 밟습나다;

then executed queries such as:

```
select 가 from 한국시 where 가 like N'%고향%'
```

which returns

가
내 고향 남쪽바다 그 파란 물 눈에 보이네; 꿈에 들 잊으리오 그 잔잔한 고향바다;

Note the "N" prefixed to the operand in the like clause. This instructs to interpret the following as Unicode characters (two bytes instead of one as in ASCII). An interesting thing about Korean Unicode is that all known syllables (고 and 향 are syllables) constructed from Korean letters (ㄱ and ㅎ are letters) have a distinct code, and text is stored as such. This is fascinating – all legitimate syllables in the Korean language have been cataloged and indexed, the entire population, complete (there are 11,172 Unicode encoded Korean syllables, and they are ordered according to Korean "alphabetization" rules – 간 precedes 만, and so forth). Because the Korean syllable is roughly in quadrants, read left to right, top to bottom, and with many variations, there is likely no expedient method of storing and searching Korean text as is done with English, with its linear alphabet. This approach works remarkably well, giving us the answer, "Yes, you can save and search Korean text in SQL." And it is very natural and intuitive. By the way, "내 고향 남쪽바다" roughly translates to, "My hometown southern sea."