

## Bilgisayar Bilimleri İçin İki Kategorizasyon Sistemi

26-Ocak-2010

Bir nesne ya da olgunun bütün içerisindeki yerinin belirlenmesi ve diğer nesne ya da olgularla ilişkilerinin betimlenmesi sürecine kategorizasyon deniyor. Kategorizasyonun sentezlemeyi artırarak öğrenmeye katkıda bulunduğunu söyleyebiliriz. Sentezleme süreci de *anlamsal bellek (semantic memory)* içerisindeki ilişkileri sağlamlaştırıyor olabilir.

İngilizce'de *categorization*, *taxonomy* ve *classification* terimleri birbirleri yerine kullanılabilir olsa da bunların etimolojileri ve kavramsal vurgulamaları arasında bazı farklılıklar göze çarpıyor. Örneğin *taxonomy* terimi daha çok (etimolojisi biyoloji ile ilgili) nesnelere ve olgulara altlık üstlük ilişkisi içerisinde sınıflandırmak için kullanılırken, *categorization* daha genel sınıflandırmalar için kullanılan bir terim. Yani *taxonomy* bir ağaç yapısını çağrıştırırken *categorization* bir graf yapısını çağrıştırıyor. *Classification* teriminin ise daha genel anlamda kullanıldığını görüyoruz. Fakat ne olursa olsun bu üç terimin aralarındaki farklar çok belirgin değil ve bu terimler birbirlerinin yerine kullanılabilir.

Bir konu hangi konunun içindedir? O konunun alt konuları nelerdir? Diğer yakın konularla ilişkileri nasıldır?.. Kategorizasyon bu sorulara yanıt oluşturuyor. Örneğin:

- C nedir?
  - Bir programlama dilidir.
- Programlama dili nedir?
  - Bir bilgisayar dilidir.
- Bilgisayar dili nedir?
  - Kurgusal bir dildir.
- Kurgusal dil nedir?

....

Peki C yalnızca bir programlama dili midir?..

Ben bu yazıda bilgisayar bilimleri için çokça karşılaşılan iki kategorizasyon sisteminden bahsedeceğim. Birincisi ACM'nin (*Association for Computing Machinery*) sistemi. ACM bilimsel makaleleri bu sisteme göre sınıflandırıyor. (Ben de ACM üyesi olarak ACM'in dijital kütüphanesinden ve buradaki makalelerden uzunca bir süredir faydalaniyorum: [www.acm.org](http://www.acm.org)) Diğer ise Wikipedia'nın kullandığı sistem. Wikipedia oldukça zengin bir ansiklopedik kaynak. Günbegün daha da zenginleşiyor.<sup>[1]</sup>

ACM kategorizasyon sistemi 4 düzeyli (yani yüksekliği 4 olan) bir ağaç yapısı biçiminde. ACM bu ağaç yapısını birkaç kez yeniledi. Ağacın son uyarlaması 98 yılında yapıldı (<http://www.acm.org/about/class/>). ACM sınıflama ağacının ilk üç düzeyi kodlanmış elemanlardan, son düzeyi ise kodlanmamış elemanlardan oluşuyor. İlk düzeydeki elemanlar harflerle temsil edilmiş. Aşağıda ilk düzey elemanları görüyorsunuz:

- A. General Literature
- B. Hardware
- C. Computer Systems Organization
- D. Software
- E. Data
- F. Theory of Computation
- G. Mathematics of Computing
- H. Information Systems
- I. Computing Methodologies
- J. Computer Applications
- K. Computing Milieux

İkinci düzey elemanlar numara içeriyor. Örneğin, "D (Software)" düğümünün alt düğümleri şöyle:

- D. Software
  - D.0 GENERAL
  - D.1 PROGRAMMING TECHNIQUES
  - D.2 SOFTWARE ENGINEERING
  - D.3 PROGRAMMING LANGUAGES
  - D.4 OPERATING SYSTEMS
  - D.m MISCELLANEOUS

İlk düzey dışındaki her düzeyin sonunda "m" kodlu bir "Miscellaneous" bölümü bulunuyor. Diğer konuların içerisine girmeyen konular bu bölüme yerleştiriliyor. Üçüncü düzey de yine numara içeren elemanlardan oluşmakta. Örneğin, "D.3 PROGRAMMING LANGUAGES" düğümünün üçüncü düzey alt düğümleri şöyle:

- D. Software
  - D.3 PROGRAMMING LANGUAGES
    - D.3.0 General
      - D.3.1 Formal Definitions and Theory
      - D.3.2 Language Classifications
      - D.3.3 Language Constructs and Features
      - D.3.4 Processors
      - D.3.m Miscellaneous

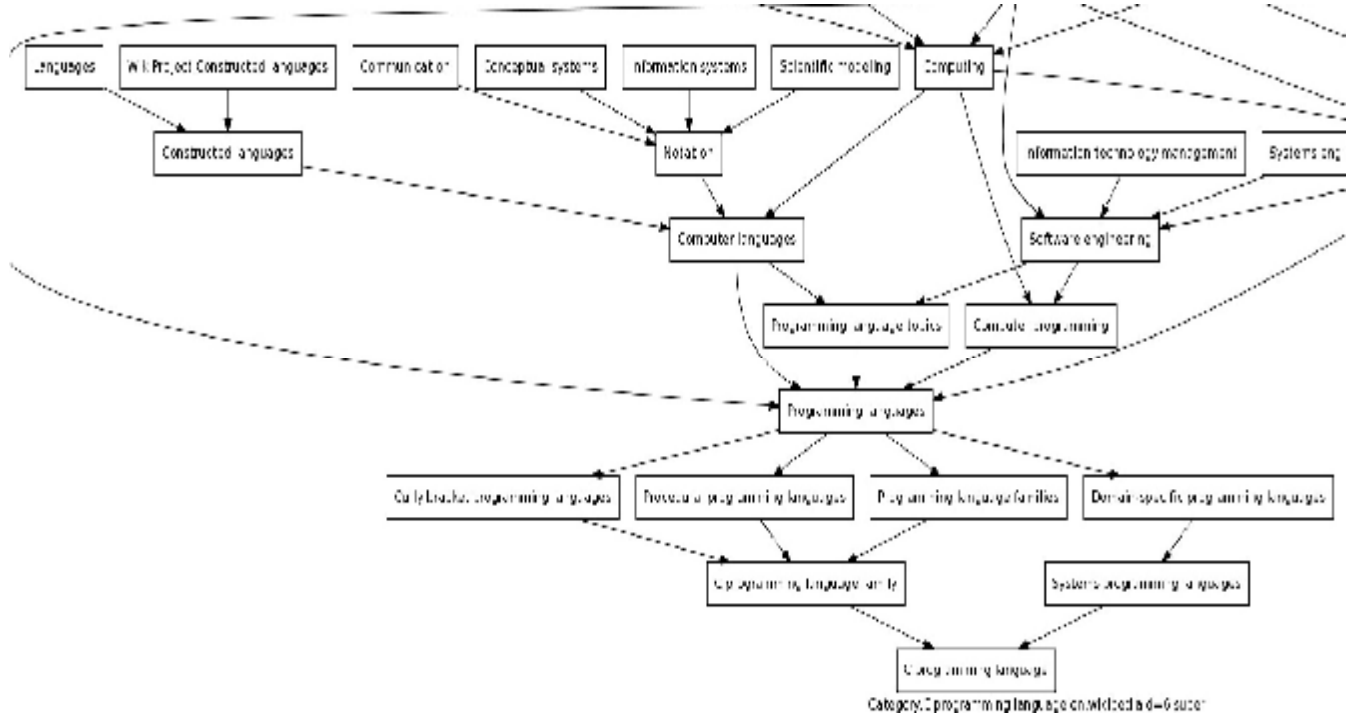
Ağacın dördüncü düzey elemanları kodlanmamış biçimde bulunuyor. Bu düzeydeki elemanlara "Implicit Subject Descriptors" ya da "Proper Noun Subject Descriptors" deniyor ve bu elemanların güncellemesi sıkça yapılıyor. Örneğin, "D.3.2 Language Classifications" düğümünün kodlanmamış elemanlarının bir bölümü şöyle:

D. Software  
D.3 PROGRAMMING LANGUAGES  
D.3.0 General  
D.3.1 Formal Definitions and Theory  
D.3.2 Language Classifications  
...  
BRL  
Bertrand  
Borland C++  
C  
C#  
C++  
CADLAN  
CDL\*  
CHILL  
CIL  
CIP  
CLEAR  
CLIST  
CLIX  
CLOS  
...

Bu durumda “C Programlama Dili”nin kategorizasyonu ACM’in sistemine göre aşağıdaki gibidir:

D. Software  
D.3 PROGRAMMING LANGUAGES  
D.3.2 Language Classifications  
C

Şimdi gelelim *Wikipedia*’ya. *Wikipedia*’nın kategorizasyon sistemi yönlü (*directed*) ve döngüsellik içerebilen bir graf (*cyclic graph*) biçiminde. Graf kategorilerden ve makalelerden oluşuyor. Bir kategorinin birden fazla üst kategorisi (*parent category*) olabiliyor. Başka bir deyişle bir kategori birden fazla kategorinin içerisinde bulunabiliyor. Makaleler (ya da *Wikipedia* terimiyle sayfalar) ise kategorilerin içerisinde. Ayrıca, makalelerin de tek bir kategori içerisinde bulunması zorunlu değil. Bir makale birden fazla kategori içerisinde bulunabiliyor. O halde şöyle özetleyebiliriz: *Wikipedia*’da bir makale bir ya da birden fazla kategori içerisindedir. O makalenin içinde bulunduğu kategoriler de bir ya da birden fazla kategori içerisinde bulunabilirler. Örneğin *Wikipedia* kategori grafının “C Programlama Dili” ile ilgili bölümü şöyle:



Algısal karmaşıklığı engellemek için grafin büyük kısmını kırtıptım. Gördüğünüz gibi "*C Programlama Dili*" kategorisi hem "*C Programlama Dili Ailesi*" hem de "*Sistem Programlama Dilleri*" kategorilerinde bulunuyor. Grafin şu kısmına dikkat ediniz:

```

Languages
├── Constructed languages
│   ├── Computer languages
│   │   ├── Programming languages
│   │   │   ├── Procedural programming languages
│   │   │   │   └── C programming language
│   │   │   ├── Programming language families
│   │   │   └── Domain-specific programming languages
│   │   └── Systems programming languages
│   │       └── C programming language
│   └── Notation
│       ├── Information technology management
│       └── Systems eng.
└── Artificially Constructed languages
    ├── Communicator
    ├── Conceptual systems
    ├── Information systems
    ├── Scientific modeling
    └── Computing

```

*Wikipedia*'da bir makaleyi açtığınızda o makalenin hangi kategoriler içerisinde bulunduğu sayfanın en altında veriliyor. Ayrıca belli bir kategorideki alt kategorilerin ve makalelerin listesi de yardımcı yazılımlarla elde edilebiliyor. Örneğin, *Wikipedia*'daki makale ve kategorilerin graflarını <http://toolserver.org/~dapete/catgraph/> sitesinden faydalanarak çizdirebilirsiniz. Ben de yukarıdaki grafi bu siteyi kullanarak oluşturdum.

[<sup>1</sup>] *Wikipedia*'nın akademik bir kaynak olarak kullanılamayacağını belirtmeliyim. Bunun pek çok nedeni var. Fakat burada ben bunlardan söz etmeyeceğim.