

**Computer Engineering Department
Bilkent University**

CS533: Information Retrieval Systems

Assignment No. 4

March 24, 2009

Due date: March 31, 2009; Tuesday, by 23:59

Please do some **MORE** annotation for the topic-based novelty detection project by following the pointers given below (in Turkish using Turkish news). The explanation provided below is the same as the one given in assignment no. 2.

Good news: Note that this time you are assigned more number of topics; however, they involve fewer number of news articles and therefore easier to annotate.

**TÜRKÇE KONU İZLEMEDE YENİLİK BULMA (NOVELTY DETECTION)
DENEY DERLEMİ HAZIRLANMASI**

TÜBİTAK destekli araştırma projemizde kullanmak ve ileride de isteyen araştırmacılarla paylaşmak üzere deney derlemi hazırlanacaktır. Bu derlemin temelini yine grubumuzun daha önceki bir çalışması olan Türkçe haberlerde yeni olay bulma ve izleme deney derlemi oluşturuyor. Sizden bu çalışmada istediğimiz daha önce yine insanlar tarafından konu izleme yönünden etiketlenmiş profilleri bir zaman çizgisi içerisinde izleyip ilk haberle başlayarak eski dökümanlarla ilgisiz, sadece ilgili ve hem ilgili hem de yenilik içeren cümleleri etiketleniz.

Profili tanımlamamız gerekirse, bir profil bir ilk haber (First Story) ve zaman içerisinde onu izleyen ilgili(Bütün haliyle döküman olarak, cümle bazında ilgisizler olabilir) haberlerden oluşuyor.

Etiketlemeyi yapacağınız sistem <http://139.179.21.199/TTracker2/index.aspx> adresindedir. Sisteme öncelikle kullanıcı girişi yapmanız gerekiyor. Kullanıcı adı, adınız ve şifresi de 20X0'li öğrenci numaranız olacak şekilde sistemde hesaplar açıldı. Çift isimliler için ilk isim kullanıldı, Mehmet Can'a Mehmet gibi. (Tek ayrıcalık, Cem Mergenci'ye Mergenci kullanıcı adı atandı).

Sisteme girdiğinizde Şekil 1'de de göreceğiniz üzere size atanmış olan profiller ekranda görünecektir. İsterseniz burada bağlantıların üzerine tıklayıp o profille ilgili bilgi edinebilirsiniz. Bir profili etiketlemek için solundaki butonu seçili hale getirip, aşağıdaki "Select" butonuna basıyoruz.

Novelty Detection**Select Profile for Novelty Detection**

- ☐ [182 - Tokyo'da Haremlik Selamlık Metro](#)
- ☐ [238 - Bill Gates türkiye'ye geldi.](#)

[Log out](#)**Şekil 1. Başarılı kullanıcı girişinden sonra karşılaşılan ekran**

Yönlendirildiğimiz sayfada(Şekil 2) seçtiğimiz profilin içindeki haberler zaman sırasıyla karşımıza çıkıyor. Burada öncelikle yapacağımız “First Story of Topic” i okumak. Bu haberle ilgili etiketleme yapılmıyor.

Bundan sonra listelenen haberleri sırayla açıyoruz(Sistem bir dökümanı tamamlamadan, diğerine geçmeye izin vermiyor. Eski her zaman dönüp bakabilirsiniz) ve cümleleri önceki haberlere ve “First Story” e göre etiketliyoruz. 3 farklı etiket var, “Not relevant”(Sarı), “Relevant but not novel”(Açık mavi) ve “Relevant and Novel”(Açık Yeşil). Her cümle öncelikle “Unprocessed” durumunda başlıyor. Eğer bir cümle profilin akışıyla ve önceki dökümanlarla alakalı değilse “Not relevant” olarak işaretliyoruz. İşaretlemeyi cümlenin üzerine tıklayarak yapıyoruz. “Unprocessed” bir cümleye bir kez tıkladığımızda “Not relevant” oluyor ve rengi sarıya dönüşüyor. Etiketler ve onlara denk gelen renkler için Şekil 3’ e bakabilirsiniz. Eğer cümle önceki dökümanlarla alakalı fakat

yeni bir bilgi içermiyorsa, cümlede bahsedilen konular önceki cümlelerde

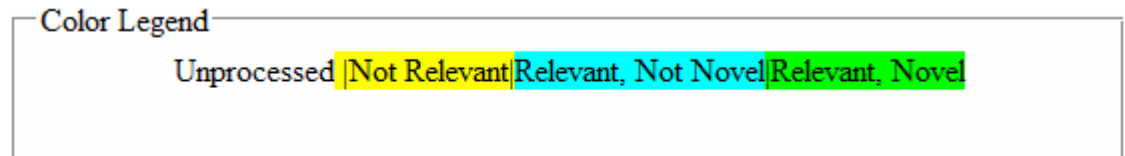


Şekil 2. Bir profil seçildikten sonra yönlendirildiğimiz ve dökümanların listelendiği sayfa

(“First story”, önceki dökümanlar veya dökümanın kendisinde önceki cümlelerde) bahsedildiyse, “Relevant but not novel” olarak işaretliyoruz. Bunun için “Unprocessed” cümleye 2 kez tıklamak yeterli(Veya “Not relevant” a bir kez). Cümle yeni bir bilgi içeriyorsa, bir kez daha tıklayıp, “Relevant and Novel” olarak işaretliyoruz.

Etiketleme sayfasında sağ tarafta gördüğünüz küçük pencerede ise bu profilde bulunan ve şimdiye kadar etiketlediğiniz dökümanlar gösteriliyor. (Şekil 4) Yani bir bilgi önceki dökümanlarda geçmiş mi diye bakmak istediğinizde, önceki dökümanları tekrar açmaktansa burayı kullanabilirsiniz. Bir dokümanla ilgili etiketlemeniz bittiğinde veya yarıda kaydedip çıkmak istediğinizde aşağıdaki “Save&Close” butonunu kullanmanız gerekiyor. Pencereyi direk kapatırsanız, yaptığınız etiketlemeler kaydedilmeyecektir.

Ancak bir dökümanı bitirdikten sonra sonraki dökümanı görüntüleyebilirsiniz. Bir sayfada 20 döküman gösteriliyor, eğer size atanan bir profil 20’den fazla döküman içeriyorsa, ”Next “butonuna basıp bir sonraki sayfaya geçebilirsiniz.Bu şekilde devam edip tüm dökümanları bitirdiğinizde ise “Evaluation Time”



Şekil 3. Cümlelerin renklerine göre ne manaya geldiklerini gösteren gösterge

menüsünden toplam o profil için ne kadar zaman harcadığınızı girmenizi istiyoruz. Bunu da yaptıktan sonra “Save&Exit” butonuna basarak çıkıp, diğer profile geçebilirsiniz.

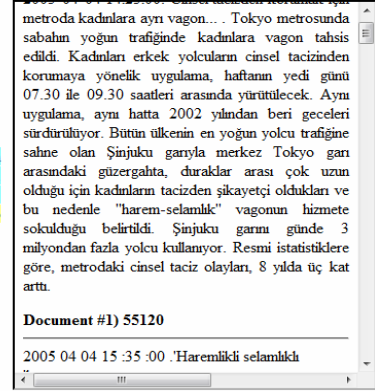
Document Position 1

2005 04 04 15 :35 :00 .Haremlikli selamlıklı 'koruma .

Color Legend

Unprocessed [Not Relevant] Relevant, Not Novel [Relevant, Novel]

Aynı uygulama ,aynı hatta 2002 yılından beri geceleri sürdürülüyor .Bütün ülkenin en yoğun yolcu trafiğine sahne olan Şinjuku garıyla merkez Tokyo garı arasındaki güzergahta ,duraklar arası çok uzun olduğu için kadınların tacizden şikayetçi oldukları ve bu nedenle "harem selamlık "vagonun hizmete sokulduğu belirtildi .Şinjuku garını günde 3 milyondan fazla yolcu kullanıyor .Resmi istatistiklere göre ,metrodaki cinsel taciz olayları ,8 yılda üç kat arttı .Tokyo Metrosu'ndaki bu uygulama Türkiye'de yapılsaydı acaba nasıl bir fırtına kopartırdı sorusunu akıllara getirdi .



Save & Close

Şekil 4. Etiketleme Sayfası

Sistemi sekme(“tab”) özelliği olan browserlar kullanmanız işinizi kolaylaştıracaktır.(Yeni dökümanları açarken) Sistem Firefox ve Internet Explorer üzerinde test edildi, başka bir browser kullanmamanızı rica ediyoruz. Bir profili tamamlamadan terkedip diğerine geçerken, açık olan profil ile ilgili pencereleri kapatarak geçmelisiniz. Diğerine geçiş yaptıktan sonra, eski açık pencerelerde işlem yapmanız sorunlara yol açacaktır.(Etiketleme ekranı için “Save & Close”, dokümanları listeleyen ekran için de “Save&Exit” butonlarını kullanınız) Kullanıcı giriş süresinin zaman aşımı 24 saattir, o yüzden sistemi açık bırakıp geri gelip devam etmeyiniz, eğer kullanıcı zamanı dolduysa yaptığınız değişiklikler kaydedilmeyebilir. Sayfaları refresh etmeyiniz, bu sizi giriş sayfasına geri götürecektir. Herhangi bir sorunla karşılaşıldığında, Cem Aksoy’a (caksoy@cs.bilkent.edu.tr) adresinden elektronik posta yoluyla ulaşabilirsiniz.