

# Sosyal Medyada İerik Analizi

Yrd.Do.Dr.Ahmet ETİNKAYA

# İçerik

- ▶ Interaktif medyanın gerektirdikleri
- ▶ Web 2.0
- ▶ Büyük Veri
- ▶ İnteraktif medya çağında içerik oluşturulması
- ▶ İnteraktif medya çağında içeriğin özel yapısı
- ▶ İnteraktif medya çağında içeriğin toplanması
- ▶ İnteraktif medya çağında içeriğin arşivlenmesi
- ▶ Uygulama örnekleri: Analyze Words, Twitter Trends

# İnteraktif medya...

- ▶ İnteraktif medya kullanıcıları sadece geleneksel medyada olduğu gibi “alıcı” konumda değildirler. İçerik değiştirebilir ve hatta üretebilirler.
- ▶ İnteraktif medya devrimi 21. yüzyılın Web 2.0 teknolojisi ile doruk noktasına ulaşmıştır.

# Web 2.0

- ▶ Web 2.0 terimi, dot-com (.com) yatırımlarının patlama yaptığı 2001 yılında internetteki değişimi tanımlamak için kullanılmıştır. Bu süreçte Wikipedia, Google ve kişisel bloglar sahneye çıktı ve internette içerik üretiminde devrim yaptılar. Web 2.0'da arayüzlerin devreye girmesiyle ortalama bir kullanıcı herhangi bir teknik bilgiye gerek duymadan içerik üretebilir oldu.
- ▶ Web 2.0'nın temel nitelikleri (Potter 2011):
  - ▶ Kişiler her türlü açık web sitesi aracılığıyla kendi çalışmalarını paylaşabilirler.
  - ▶ Kişiler her türlü siteye erişebilirler, istediklerini kullanabilirler, kendi mesajlarını oluşturabilirler ve mesajlarını herkesin görmesini sağlayabilirler.
  - ▶ Açık paylaşım özelliği yaratıcı aktivitelerde büyük artışa yol açmıştır.

# Devasa Online İerik...

- ▶ Web 2.0 ve yksek hızlı kablosuz internet eriřimi imkanı byk miktarlarda ierik retilmesine yol amıřtır.
- ▶ Emsalsiz hacimde ve sınırsız eriřimde kiřisel, kiřilerarası, grup, kurumsal, kitlesele ve diđer her trl mesaj
- ▶ Yazılı ieriđin yanısıra ok fazla miktarda iřitsel ve grsel ierik de mevcut.
- ▶ Youtube'nın 2015 raporuna gre: her dakika ortalama 300 saatlik toplam video ykleme ve 75 lkeden toplam 1 milyar kullanıcı.

# Interaktif medyanın içerik analizine getirdiđi zorluklar

- ▶ Geleneksel medyada içerik belirli formlardaydı: Gazete sayfası, televizyon dizisi gibi. Ancak, interaktif medya dinamik ve akışkan bir yapıdadır.
- ▶ Intertaktif medyaya has bu yeni yapı, bunun üzerinde içerik analizi yapılmasını ve hatta bu medyanın tespit edilmesini zorlaştırmaktadır. (örneğin bir kişinin Facebook'ta neyi paylaştığını sadece metin madenciliđi ile tespit etmeniz mümkün değildir)
- ▶ Web sayfaları ve diđer içerikler her an güncellenebilirler. Örneđin Snapchat ve Periscope gündelik ve canlı iletişimi maksimize etmek üzere oluşturulmuş platformlardır. Bu, içerik analizi metodunda ciddi zorluklara yol açmaktadır.
- ▶ Online içerikteki aşırı artışla birlikte araştırmacılar kullanıcı davranışlarının şekillendirdiđi devasa veri kümeleri ile karşı karşıya kaldılar. Sosyal medya platformları, web siteleri ve mobil medya uygulamalarında insan davranışı tarafından oluşturulmuş, biriktirilmiş ve arşivlenmiş bu bilgi kümelerine “Büyük Veri” adı verilmiştir.

# Büyük Veri

- ▶ İnsanlar tarafından ele alınıp analiz edilemeyecek ve hatta tümüyle anlaşılamayacak derecede büyük hacimli ve kompleks veri setleridir.
- ▶ Toplanması ve analiz edilmesi için bilgisayarların kullanılması gerekir. Bu yüzden datanın tamamen anlaşılıp hazmedilmesi insanlar için mümkün değildir. Sadece data ile ilgili genel özet bilgilere varılabilir. Sonuçlar genelde soyuttur ve kişiler tarafından tecrübe ile kanıtlanmaz.
- ▶ Farklı disiplinlerden bilim insanları Büyük Veri ile ilgili endişelerini dile getirmişler ve Büyük Veri analizlerinin, bilimsel araştırmanın “temsili örnekleme” ve “tümünden gelim” kurallarını gözardı edebileceğinden söz etmişlerdir. (çünkü, Büyük Veri analizleri genellikle tümevarım niteliğindedir ve veriden bağımsız değildir.)
- ▶ Bütün bu endişelerden dolayı bilgisayar destekli metin analizi yöntemleri ile manuel (kişisel kodlama) yöntemlerin birlikte kullanılmasıyla; kişisel kodlama ile sağlanacak sistematik kesinlik ve içerik farkındalığı avantajları ile büyük verinin büyük-çaplı kapasitesi ve bilgisayar destekli yöntemlerin etkinliği biraraya getirilmiş olur.

# Sosyal Medya İerik Analizi alıřmalarından rnekler

- ▶ Anneler iin aılmıř Facebook sayfaları (Kaufmann & Buckner, 2014)
- ▶ Twitter ve Weibo'nun kltrlerarası bir ierik analizi alıřması (Ma, 2013)
- ▶ Twitter'da ocukluk obezitesi (Harris et. al., 2014)
- ▶ Tunus ve Mısır devrimleri sırasında atılan Twitter mesajları (Hermida, Lewis & Zamith, 2014)
- ▶ YouTube'da direkt kullanııcıya ynelik farmasotik reklamlar (Muncy, Iyer & Eastman, 2014)
- ▶ Fortune 500'n Facebook sayfaları (McCorkindale, 2010)
- ▶ COPD hastalıęı ile ilgili bir saęlık bilgisi kaynaęı olarak Pinterest (Paige et al., 2015)



# İnteraktif medya çağında içerik oluşturulması

- ▶ Mesaj kaynağının ve mesajın ayrıştırılabilmesi geleneksel medyadaki gibi kolay ve kesin değildir. İnteraktif medya çağında “seyirciler/dinleyiciler” kendi medya ortamlarını şekillendirme kabiliyetine sahiptirler.
- ▶ Kişilerin, örneğin, bir web sitesinde bilgiyle etkileşimleri sitenin neresinde gezindiklerine, ne kadar sık tıkladıklarına ve geçmiş online davranışlarına bağlıdır (mesela geçmiş web gezintisi davranışları yayınlanacak reklamları belirler).
- ▶ Web 2.0’ın gelişimiyle kullanıcıların içerik geliştirebilmesi ve silebilmesi analiz edilecek yığınla medya ürününün doğmasına sebep olmuştur.

# İnteraktif medya çağında içerik oluşturulması

- ▶ Tek başına Facebook 1,5 milyondan fazla aktif kullanıcıya sahiptir ve bu kullanıcıların her birinin kişisel sayfası bulunur. Bu her bir kişisel sayfada analiz edilebilecek birden fazla öge bulunur: duvar yazıları, fotoğraf albümleri, görüntüler, başka web sitelerinden alıntılanan paylaşımlar, kişisel bilgiler ve çok daha fazlası.
- ▶ Araştırmacıların karşılaştığı bir diğer zorluk kişilerin kendi özgün içerikleri ile diğer web sitelerinden alıntıladıkları içeriklerin ayrıştırılmasıdır. Multimedya içerik oluşturma araçlarının gelişmesiyle birlikte bir fotoğrafın, videonun veya herhangi başka bir multimedya içeriğin hesabın sahibi tarafından oluşturulup oluşturulmadığını tespit etmek gittikçe zorlaşmaktadır. Kullanıcılar çoğunlukla diğer kaynaklardan aldıkları içerikleri link olmaksızın paylaşmaktadırlar.

# İnteraktif medya çağında içeriğin özel yapısı

- ▶ 1. Kullanıcı tarafından oluşturulan içerik (User-generated content - UGC)
  - ▶ Örnekler: Facebook profilleri, Instagram resimleri ve videoları, emailler, Twitter, Youtube
  - ▶ UGC'lerin büyük bir kısmı kişilerarasıdır; Facebook ve email gibi. Bir kısmı kişiseldir ancak daha geniş bir kitle tarafından erişimi vardır; sosyal ağ aktiviteleri gibi. Bazı UGC'ler de "kitlesele"dir; Wikiler ve Youtube videoları gibi.
- ▶ 2. Kullanıcı tarafından seçilen (veya toplanan) içerik (User-selected content - USC)
  - ▶ Örnekler: Bir TV programının Youtube'da paylaşılması, BitTorrent'te müzik paylaşılması, YouTube'daki bir videonun Facebook veya Twitter'da paylaşılması.
  - ▶ USC'ler bir kişinin mesaj seçimlerini göstermesi bakımından önemlidir.
  - ▶ USC'ler aynı zamanda insanların online nasıl gezindiklerinin ve online içerikleri nasıl anlamlandırdıklarının bir göstergesidir.

# Öneriler

- ▶ Kullanıcı tarafından oluşturulan içeriğin inceleneceği durumlarda, web sayfalarının sadece belirli bölümlerinin kullanıcılar tarafından kontrol edilebileceği göz önünde bulundurulmalıdır. Örnek 1: YouTube’da kullanıcılar video paylaşabilir ancak “ilgili linkler” bölümü kullanıcı tarafından oluşturulamaz. Örnek 2: Facebook kullanıcıları kendi sayfalarındaki veya profillerindeki reklamları kontrol edemezler.
- ▶ İçerik analizinde ayrıca kullanıcının sayfasında kendisinin paylaştığı içerik ile başkasının paylaştığı içerik ayrıştırılmalıdır (Facebook Zaman tüneli veya Youtube yorumlar bölümlerinde olduğu gibi)
- ▶ İncelenmek istenen içeriğin “kullanıcı tarafından oluşturulan içerik” mi “kullanıcı tarafından seçilen içerik” mi olduğuna dikkat edilmelidir. Örneğin bir sosyal medya sitesinde paylaşılan mesajın “kaynak”ı UGC’de hesabın sahibi iken USC’de “kaynak” hesabın sahibi değildir.

# İnteraktif medya çağında içeriğin toplanması

- ▶ İnternetin yayılması sebebiyle içerik toplanması hiç olmadığı kadar kolaylaşmıştır.
- ▶ Örnekler: LexisNexis, YouTube, Netflix, DVRlar, iTunes, Sosyal medya içerikleri, Evernote
- ▶ Sosyal ağ sitelerinden içerik almaktaki temel bir zorluk yaygın olarak kullanılan güvenlik araçlarıdır. Sosyal paylaşım ağları kişisel bilgilerin sadece belirlenmiş bir arkadaş çevresiyle karşılıklı olarak paylaşılmasına imkan tanır. Bu yüzden, böyle bir popülasyondan örnekleme yapmak zordur. Kişisel güvenlik problemini aşmak için ya kullanıcıların onayı alınmalı ya da halka açık profiller üzerinden araştırma yapılmalıdır.
- ▶ Sosyal ağ sitelerinden örnekleme yapmak için öncelikle sitenin kendi gelişmiş tarama aracı kullanılabilir.

# İnteraktif medya çağında içeriğin toplanması

- ▶ 2006'da kurulan Twitter şu anda 320 milyondan fazla aktif kullanıcıya sahiptir. Bu kullanıcılar günde 500 milyondan fazla Tweet üretmektedirler.
- ▶ Twitter'da örnekleme yapılmasında hashtagler kullanılabilir. @ cevaplar da Twitter'da ayırıcı bir özelliktir. Ayrıca, Twitter'ın kendi sunduğu API'ler aracılığıyla Twitter örnekleme yapılabilir. Ancak, bunun için belli düzeyde teknik bilgiye ihtiyaç vardır.

# İnteraktif medya çağında içeriğin arşivlenmesi

- ▶ Web sitelerinin arşivlenmesi ilk çıktıkları HTML formdaki yapılarına kıyasla şu anda daha zordur. Şimdiki sitelerde “sayfayı kaydet” fonksiyonunun kullanılamayacağı birçok farklı eleman bulunmaktadır.
- ▶ Ancak, web sitelerinin arşivlenmesinde birçok yöntem kullanılabilir.
- ▶ Örneğin, Shelton ve Skalski (2014) Facebook sayfalarının arşivlenmesi için Adobe Acrobat kullanmış ve sayfaların anlık görüntüsünü içeren PDF dosyaları oluşturmuşlardır.

# İnteraktif medya çağında içeriğin arşivlenmesi

- ▶ Sosyal medya mesajlarının arşivlenmesinde Evernote da kullanışlıdır. Evernote çok farklı türde mesajların toplanmasını kolaylaştıran bir yazılımdır ve içeriğin yanısıra metinle ilgili diğer bilgileri de kaydedebilir. Mesajın bir görüntüsünün veya resminin alınmasını ve Evernote'un bulut serverına ve kullanıcının masaüstüne kaydedilmesini sağlar. Mesajı arşivleyen kullanıcı mesajı diğer araştırmacıların kullanımına da açılabilir. Evernote optik karakter tanıma ile resimdeki yazıları tarama yapılabilecek metne dönüştürür.
- ▶ Bir web sitesinin tümünün kaydedilmesi için geliştirilmiş yazılım araçları da bulunmaktadır. HTTrack Website Copier bunlardan biridir.



# İçerik Analizi gibi...

- ▶ Online içerik ve mesajlardaki aşırı artışla birlikte içeriğin analiz edilmesi (veya içerik analizinin yakın akrabası olan metin matenciliği) büyük bir iş sahası haline gelmiştir.
- ▶ Radian6 ve Sprout Social gibi sosyal medya analiz yazılımları Twitter, Facebook, Youtube, bloglar, online topluluklar ve diğer benzer sitelerden içerik verisi çekebilmektedirler. Bu programlar sosyal bilimler alanındaki araştırmacılardan ziyade ticari amaçlı olarak sunulmaktadır. Ancak, araştırmacılar için de kullanışlı olabilirler.
- ▶ Ticari amaçlı pahalı yazılımların yanısıra içerik analizine yönelik ücretsiz yazılımlar da artmaya başlamıştır.
- ▶ Bu yazılımların yaptığı peki gerçekten içerik analizi midir? Çoğu durumda bu tür programlar metin madenciliği yapan özelleşmiş uygulamalardır. Ancak, bu yazılımların bazıları artık daha oturmuş kodlama yapıları kullanmaya başlamışlardır. Akademik anlamda tam olarak içerik analizi olup olmaması bir tarafa, bu uygulamalar, çok büyük hacimlerdeki mesajların hızlıca analiz edilmesinde bilgisayar programcılığının gücünü göstermektedir.

# Uygulamalardan örnekler

- Bir Twitter kullanıcısı ile ilgili kişilik analizi oluşturulmasını sağlar. Bu kişilik analizi duygusal stil, sosyal stil ve düşünme stili analizlerini içerir.

## ANALYZE WORDS

**What does AnalyzeWords do?**  
AnalyzeWords helps reveal your personality by looking at how you use words. It is based on good scientific research connecting word use to who people are. So go to town - enter your Twitter name or the handles of friends, lovers, or Hollywood celebrities to learn about their emotions, social styles, and the ways they think.

Twitter Handle: @katyperry  The Development Team  
The Science Behind AnalyzeWords

Analyze Friend: - select -  http://analyzewords.com/index.php?handle=katyperry

---

Share:

### Analysis of tweets from katyperry (596 most recent words - 8th May, 2017)

#### Emotional Style

Upbeat (Average)	44	<div style="width: 44%;"></div>
Worried (Very low)	20	<div style="width: 20%;"></div>
Angry (Low)	27	<div style="width: 27%;"></div>
Depressed (Average)	54	<div style="width: 54%;"></div>

#### Social Style

Plugged In (Average)	51	<div style="width: 51%;"></div>
Personable (Low)	32	<div style="width: 32%;"></div>
Arrogant/Distant (Low)	30	<div style="width: 30%;"></div>
Spacy/Valley girl (Average)	55	<div style="width: 55%;"></div>

#### Thinking Style

Analytic (Low)	28	<div style="width: 28%;"></div>
Sensory (High)	64	<div style="width: 64%;"></div>
In-the-moment (Average)	45	<div style="width: 45%;"></div>

**Tweets analyzed from katyperry:**