

Dođal Dilden Biçimsel Dile Dönüşüm Üzerine Bir Yöntem

Nihal Yağmur Aydın

Boğaziçi Üniversitesi

Taslak

- Amaçlar
- Önceki Çalışmalar
- Sistem Dizaynı
- Yöntemler
- Tartışma ve İleri Çalışmalar

Amaçlar

- Doğal dilin makineler tarafından anlaşılması
- Biçimsel doğrulama
- Modelleri test etme
- Otomatikleştirme
- Zaman ve para tasarrufu

Önceki Çalışmalar

Metin madenciliğinde kullanılan özellik çıkarma yöntemleri:

- 1) Morfolojik Analiz (MA) : kelimelerin köklerine ayrılması
- 2) Sözdizimsel Analiz (SA) : gramatik bilgi alma işlemi
- 3) Anlamsal Analiz (Irfan ve diğ., 2015)

Sistem Tasarımı

Girdi: Doğal dildeki güvenlik politikası dokümanı

- 1) Ön İşleme
- 2) Sözdizimsel Analiz
- 3) Mantıksal Çıkarım
- 4) Biçimsel Dile Dönüşüm

Çıktı: Biçimsel dildeki doküman

Kümeleme Yöntemi

Varsayım: Roller önceden tanımlanmış olmalıdır.

- Güvenlik politikası dokümanındaki cümleler, rollerin özne olması durumunda kümelenebilir.
- Bu sayede mantıksal çıkarım işinde kolaylık sağlanmış olur.

Sözdizimsel Analiz

- Sözcük türü etiketleme (POS tagging)

Girdi:

System administrators must maintain formal records or logs of all exchanges of data and software between organizations.

Çıktı:

System/NNP Administrator/NNP must/MD maintain/VB formal/JJ records/NNS or/CC logs/NNS of/IN all/DT exchanges/NNS of/IN data/NNS and/CC software/NN between/IN organizations/NNS ./.

Sözdizimsel Analiz

Ayrıştırma Ağacı (Parse tree)

- Dönüşüm işleminde işe yaramayacak bilgilerin çıkarılması için kullanılabilir.
- Özne (isim) -fiil-nesne(isim) sıralamasının cümle içindeki tespiti, ayrıştırma ağacı ve sözcük türü etiketlemenin birleşimi ile mümkün olabilir.

Metinde Mantıksal Çıkarım

- **Tümdengelimli Usavurma**

Tümdengelim yönteminde, genellemelerden yola çıkarak bileşenlere ayrıca ulaşmak mümkündür (Huth & Ryan,2004).

Girdi:

“IT staff should report any security incidence to the System Administrator.”

Çıktı:

IT Staff = Application Developers V Database Admin V Web Admin

Metinde Mantıksal Çıkarım

- **Tümevarımlı Usavarma**

Tümevarımlı usavurma yönteminde, bileşenlerden genelleme yapılmaktadır. Güvenlik politikası metinlerinde tümevarım yönteminin kullanımı ise, “all staff” adı altında geçmeyen, ancak, farklı roller için tekrarlanan aynı cümleleri gruplandırmakla mümkündür.

Örnek kümeler:

security manager (güvenlik yöneticisi), risk manager (risk yöneticisi) ve system manager (sistem yöneticisi) .

Yöntem:

Kosinüs benzerliği ile cümleler arasındaki benzerlik oranlarından faydalanarak kümelerin oluşturulması, ardından öznelerden genelleme yapılması.

Doğal Dilden Birinci Derece Mantığa Dönüşüm

- fiil(özne, nesne)

Girdi: “System Administrator should assist the Security Manager about the implementation of ISPD and support the employees about the security issues.”

Çıktı: $\exists x (\text{assist}(S,x)) \wedge \forall y (\text{support}(S,y))$

System Admin=S,

Security Manager=x,

Employee=y ise.

Birinci Derece Mantıktan Biçimsel Dile Dönüşüm

- 1. derece mantıktan Z biçimsel diline dönüşüm.

Girdi: $\exists x (\text{assist}(S,x)) \wedge \forall y (\text{support}(S,y))$

Çıktı: $(\exists x : S \bullet \text{assist}(x)) \wedge (\forall y : S \bullet \text{support}(y))$

Tartışma ve İleri Çalışmalar

- Morfolojik analiz yoluyla bilgi çıkarımı
- Cümleler arasındaki anlamsal bağların incelenmesi
- Prolog gibi ortamlarda yeni dilbilgisi yapılarının tanımlanması
- Cümleler üzerinden kural çıkarma işlemi

Kaynaklar

- [1] Irfan R. , King C.K , Grages D. , Ewen S. , Khan S.U. , Madani S.A. , Kolodziej, Wang L., Chen D. , Rayes A. , Tziritas N. , Xu C. , Zomaya , A.Y. , Alzahrani A.Z. & Li H. (2015). A Survey on Text Mining in Social Networks. *The Knowledge Engineering Review*, 30(2), pp.157–170. Cambridge Press.
- [2] Spivey, J. M., & Abrial, J. R. (1992). *The Z notation*, p. 90. Hemel Hempstead: Prentice Hall.