

Ontology-Based Named Entity Recognizer for Behavioral Health

Reconocedor de entidades para la salud basado en ontologías

Resumen. - Los reconocedores de nombres de entidades son una importante de los sistemas de extracción de información en tareas de anotación. Aunque el progreso sustancial está siendo hecho en las entidades nombradas independientes del dominio, hay una necesidad de reconocer entidades para aplicaciones en un dominio específico con el fin de extraer conceptos relevantes. Debido al crecimiento de las aplicaciones de salud para mejorar las epidemias de enfermedades (sobrepeso, falta de ejercicio, alcohol, drogas, etc) enfocan su investigación en el dominio del cambio del comportamiento, concretamente el cambio de estilo de vida. Para mejorar su base de conocimiento no hay un reconocedor diseñado para el dominio de los cambios en el estilo de vida. Describen el diseño de una ontología para el comportamiento de la salud basado en que han desarrollado un aumento de NER con recursos léxicos. El sistema NER etiqueta automáticamente palabras y frases con etiquetas relevantes del estilo de vida como por ejemplo comida sana o insana, potencialmente peligroso, droga, tabaco, etc. Además discuten cómo los sistemas de ontología pueden crecer con estos reconocedores usando razonadores semánticos.

1 INTRODUCCIÓN

El estilo de vida actual nos está llevando a tener enfermedades bastante graves, pasando por sobrepeso, obesidad, diabetes tipo 2 e incluso cánceres que se podrían evitar. Debería existir una mejor medicina preventiva y las aplicaciones de salud pueden ayudar a conseguirlo.

El problema que ven es que las entidades crecen y trabajar en ellas supone bastante dinero y es un problema añadido.

Su sistema categoriza en 5 categorías lo que encuentra.

2 OTROS TRABAJOS

Los reconocedores de entidades han sido usados en otros dominios como biología o inteligencia empresarial. En el dominio del comida también diseñaron un reconocedor en 2012, recomendando productos y ayudando online a los usuarios. Se clasificaba la comida en saludable y no saludable. Este sistema era dependiente del contexto y tenía incluidas incluso alergias, con lo que requería cierta información del usuario.

Usaron WordNet y OWL y su objetivo principal era desarrollar un sistema de diálogo para el dominio del cambio del estilo de vida usando el reconocedor. Reconociendo y clasificando entidades en el dominio específico se cubría el primer paso. Teniendo una ontología OWL introducir posibilidades adicionales para las entidades reconocidas usando razonadores para clasificar en categorías de importancia.

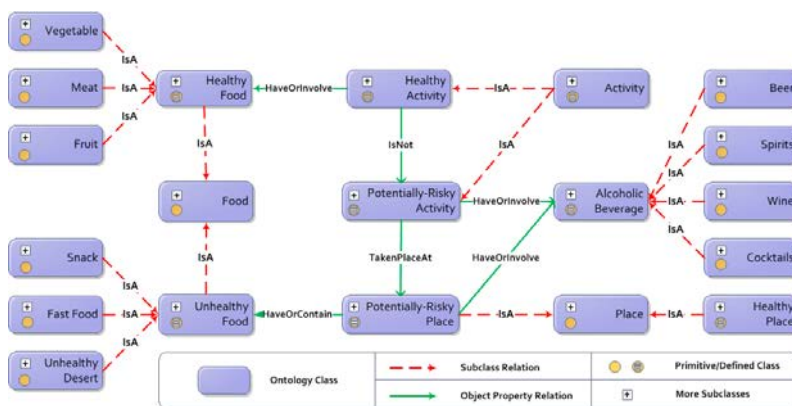
3 ENFOQUE

Diseño de ontología: Diseñaron la ontología con **Protégé** en **OWL 2**. 60 clases, 490 individuales y 20 propiedades de objetos. Saludable, insaludable y potencialmente de riesgo. Para las comidas tenían en cuenta:

- Contenido de azúcar y calorías del alimento.
- Contenido de grasas
- Aditivos artificiales y procesado industrial.

Por ejemplo, el brócoli era una instancia de la clase vegetal que era subclase de comida saludable y el chocolate era una instancia de una clase snack que era subclase de comida no saludable.

Otro término saludable era **actividad** e insaludable **sedentarismo**, etc.



WordNet y Distancia Wordnet

Utilizaron esta base de datos para hiperónimos e hipónimos, así fueron capaces de relacionar mejor la semántica. Usaron **RiTa.WordNet**

Arquitectura

Basado en un Part-Of-Speech Tagger.

El algoritmo de etiquetado era el siguiente:

Algorithm 1 Tagger

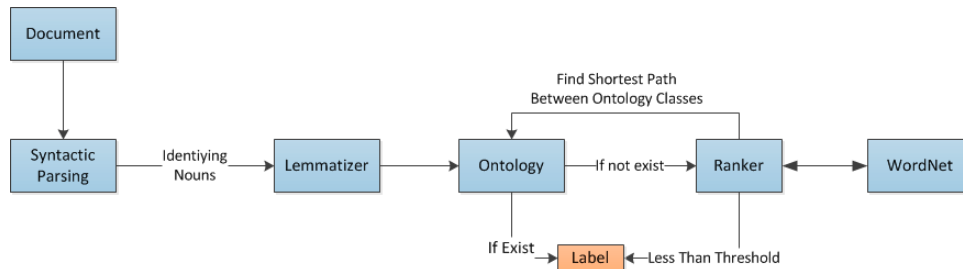
```

if Is parameter noun(pn) a class in ontology then
    Tag the noun with super class of corresponding class
else if Is pn individual in ontology then
    1.Find class Of the individual
    2.Find super class of individual's class
    Tag the pn with the super class
else
    1.Compare minimum distance between noun and Ontology
    classes by using Distance algorithm (See Algorithm
    2)
    2.Select the class with shortest path to the noun
    3.If Selected class is first level and distance is less than
    higher-threshold
    return it as Tag
    4.If selected class is not first level and distance is less
    than lower-threshold
    return it as Tag
    Otherwise do not tag
  
```

end if

4 EVALUACIÓN Y DATOS

Recogieron por ejemplo recetas de internet leyendo 88 frases y 220 entidades y comenzaron a testear el reconocimiento de los tipos de comida según lo que leían en las recetas. Si una entidad reconocía comida con sin etiquetar en saludable o no saludable, no contaba como correcta.



Algoritmo de distancia:

Algorithm 2 Distance Algorithm

Locate the common parent of the two lemmas by checking each sense of each lemma

if No common parent found then

return 1

else

1. Calculate min distance to common parent (the shortest path from either lemma to common parent)

2. Calculate distance from common parent to root (length of the path from common parent to the root of WordNet ontology)

3. Calculate and return the normalized distance to common parent as:

$(\text{minDistToCommonParent} / (\text{distFromCommonParentToRoot} + \text{minDistToCommonParent}))$

end if

5 CONCLUSIONES Y FUTUROS TRABAJOS

Diseñaron el reconocedor, marcaron las diferencias entre el reconocedor y el dominio específico. Crearon un reconocedor de entidades, identificaron posibles casos de uso para su ontología y para extender la ontología usaron Wordnet calculando distancias de conjuntos de hiperónimos.