

EtymologicalRelations

Procesamiento de datos mediante un motor de derivación lógico.

Instituto Tecnológico de Costa Rica.

Escuela de Ingeniería en Computación.

Curso de Inteligencia Artificial, I Semestre del 2018.

Instalación

Dependencias

Es necesario contar con `Python 3` para poder realizar la instalación. Además, se requieren algunas dependencias que se pueden instalar utilizando `pip`:

1. Se abre un terminal con permisos de administrador
2. Se ejecuta el siguiente comando:

```
sudo pip3 install pyDatalog
```

Instalación desde el código fuente

1. Se debe abrir una terminal con permisos de administrador
2. Se debe navegar hasta la ruta del código fuente, al nivel que se encuentre el archivo `setup.py`
3. Se ejecuta el siguiente comando:

```
sudo python3 setup.py install
```

Instalación utilizando pip

1. Se debe abrir una terminal con permisos de administrador
2. Se ejecuta el siguiente comando:

```
sudo pip3 install tec.ic.ia.p2.g03
```

Manual de Usuario

Operación 1

Descripción: Determinar si dos palabras son heman@s

Lo que pretende esta funcionalidad es que a partir de dos palabras (**X, Y**) proporcionadas por el usuario, el sistema brinde una respuesta indicando si la palabra **X** está relacionada con la palabra **Y** o no.

Entonces, suponiendo que el usuario proporciona la palabra '**X = hijo1**' y '**Y = hijo2**,' el sistema verifica en la base de conocimiento para ver si existe un hecho que cumpla la relación donde **X** and **Y** tienen un **padre en común**.

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar la palabra1 y la palabra2 en los campos de texto respectivos. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Operaciones entre palabras			
Palabra X	Palabra Y	Resultado	
hijo1	hijo2	Si	Grado primas
Es hija X de Y	Es tía X de Y	Son hermanas	Son primas

Note que la salida es simplemente un **Si/No** indicando si la palabra **X** se encuentra relacionada a la palabra **Y** bajo el concepto de herman@s.

Operación 2

Descripción: Determinar si dos palabras son prim@s

Lo que pretende esta funcionalidad es que a partir de dos palabras (**X, Y**) proporcionadas por el usuario, el sistema brinde una respuesta indicando si la palabra **X** está relacionada con la palabra **Y** o no.

Entonces, suponiendo que el usuario proporciona la palabra '**X = primo1**' y '**Y = primo2**,' el sistema verifica en la base de conocimiento para ver si existe un hecho que cumpla la relación donde **X** and **Y** tienen un **ancestro en común**.

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar la palabra1 y la palabra2 en los campos de texto respectivos. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Operaciones entre palabras			
Palabra X	Palabra Y	Resultado	
primo1	primo2	Si	Grado primas
Es hija X de Y	Es tía X de Y	Son hermanas	Son primas

Note que la salida es simplemente un **Si/No** indicando si la palabra **X** se encuentra relacionada a la palabra **Y** bajo el concepto de prim@s.

Operación 3

Descripción: Determinar si una palabra es hij@ de otra

Lo que pretende esta funcionalidad es que a partir de dos palabras (**X, Y**) proporcionadas por el usuario, el sistema brinde una respuesta indicando si la palabra **X** está relacionada con la palabra **Y** o no.

Entonces, suponiendo que el usuario proporciona la palabra '**X = padre**' y '**Y = hijo**,' el sistema verifica en la base de conocimiento para ver si existe un hecho que cumpla la relación donde Y tiene como padre a la palabra X.

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar la palabra1 y la palabra2 en los campos de texto respectivos. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Operaciones entre palabras			
Palabra X	Palabra Y	Resultado	
padre	hijo	Si	Grado primas
Es hija X de Y		Es tía X de Y	Son hermanas
			Son primas

Note que la salida es simplemente un **Si/No** indicando si la palabra **X** se encuentra relacionada a la palabra **Y** bajo el concepto de hij@.

Operación 4

Descripción: Determinar si una palabra es ti@

Lo que pretende esta funcionalidad es que a partir de dos palabras (**X, Y**) proporcionadas por el usuario, el sistema brinde una respuesta indicando si la palabra **X** está relacionada con la palabra **Y** o no.

Entonces, suponiendo que el usuario proporciona la palabra '**X = tío**' y '**Y = sobrino**,' el sistema verifica en la base de conocimiento para ver si existe un hecho que cumpla la relación donde existe un primo P tal que todos los ancestros T de P que no sean los ancestros directos de Y son tíos de Y.

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar la palabra1 y la palabra2 en los campos de texto respectivos. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Operaciones entre palabras			
Palabra X	Palabra Y	Resultado	
tío	sobrino	Si	Grado primas
Es hija X de Y		Es tía X de Y	Son hermanas
			Son primas

Note que la salida es simplemente un **Si/No** indicando si la palabra **X** se encuentra relacionada a la palabra **Y** bajo el concepto de ti@.

Operación 5

Descripción: Determinar si son prim@s y en qué grado.

Lo que pretende esta funcionalidad es que a partir de dos palabras (**X, Y**) proporcionadas por el usuario, el sistema brinde una respuesta indicando si la palabra **X** está relacionada con la palabra **Y** o no, bajo la relación de primos.

Entonces, suponiendo que el usuario proporciona la palabra '**X = primo1**' y '**Y = primo2**,' el sistema verifica en la base de conocimiento para ver si existe un hecho que cumpla la relación donde Y tiene un ti@ Z, y X tiene un ti@ W, tal que Z y W son herman@s en algún nivel (N). Entonces se cumplirá que la palabra X es prima de la palabra Y en el nivel N.

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar la palabra1 y la palabra2 en los campos de texto respectivos. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Palabra X	Palabra Y	Resultado
primo1	primo2	Primos 2°

Grado primas

Es hija X de Y Es tía X de Y Son hermanas Son primas

Operación 6

Descripción: Determinar si una palabra está relacionada con un idioma (Si / No)

Lo que pretende esta funcionalidad es que a partir de una palabra y un idioma proporcionados por el usuario, el sistema brinde una respuesta indicando si la palabra está relacionada con el lenguaje o no.

Entonces, suponiendo que el usuario proporciona la palabra '**padre**' y el idioma '**spa**,' el sistema verifica en la base de conocimiento para ver si existe un hecho que cumpla la relación, ya sea que el *idioma:palabra* se encuentre al lado derivado o al contrario(generator).

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar la palabra y el idioma en los campos de texto respectivos. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Idioma D	Palabra P	Resultado
spa	padre	SI

P se relaciona con D

Palabras originadas por P en D

Idiomas relacionados con P

Note que la salida es simplemente un **Si/No** indicando si la palabra se encuentra relacionada al lenguaje proporcionado.

Operación 7

Descripción: Obtener el conjunto de todas las palabras en un idioma originadas por una palabra específica

Lo que pretende esta funcionalidad es que a partir de una palabra y un idioma proporcionados por el usuario, el sistema brinde una respuesta indicando todas las palabras que son originadas a partir de la ingresada en la interfaz y además que dichas palabras de encuentren en el idioma indicado.

Entonces, suponiendo que el usuario proporciona la palabra base: **'-lik'** y el idioma: **'afr,'** el sistema procede a verificar en la base de conocimiento para ver si existen hechos que cumpla la relación donde la palabra generadora sea **'-lik'** y que el lenguaje de la palabra generada sea **'afr'**.

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar la palabra y el idioma en los campos de texto respectivos. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Idioma D	Palabra P	Resultado
afr	-lik	OK

wetenskaplik
tydelik
persoonlik
wetlik

P se relaciona con D

Palabras originadas por P en D

Idiomas relacionados con P

Note que la salida difiere de la operación anterior, ya que en esta no es simplemente un valor booleano, sino que se espera una lista de idiomas, y de no encontrar ningún hecho simplemente despliega una lista vacía de elementos.

Operación 8

Descripción: Listar los idiomas relacionados con una palabra

Lo que pretende esta funcionalidad es que a partir de una palabra proporcionada por el usuario, el sistema brinde una respuesta indicando todas los lenguajes que están relacionados de manera directa con la misma. Ya sea que la palabra esté escrita en un idioma o que sea generada de otra palabra en un idioma determinado. En síntesis, la respuesta se compone de los lenguajes originarios y de los generados, donde exista la palabra especificada en la interfaz.

Entonces, suponiendo que el usuario proporciona la palabra base: **'bruin'** el sistema procede a verificar en la base de conocimiento para ver si existen hechos que cumpla la relación donde la palabra generadora/generada sea **'bruin'**, entonces de cumplirse, se toman en cuenta los lenguajes que incorpora el hecho del KB.

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar la palabra en el campo de texto respectivo. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Operaciones palabra e idioma		
Idioma D	Palabra P	Resultado
No aplica	bruin	OK
nld afr		

P se relaciona con D

Palabras originadas por P en D

Idiomas relacionados con P

Note que en la salida se espera una lista de idiomas, y de no encontrar ningún hecho simplemente despliega una lista vacía.

Operación 9

Descripción: Contar todas las palabras comunes entre dos idiomas

Lo que pretende esta funcionalidad es que a partir de dos idiomas (**X, Y**) proporcionados por el usuario, el sistema brinde una respuesta indicando la cantidad de palabras que se encuentran relacionadas de manera común entre los lenguajes (**X and Y**).

Entonces, suponiendo que el usuario proporciona el lenguaje '**X = eng**' y '**Y = afr**,' el sistema verifica en la base de conocimiento para ver si existe un hecho que cumpla la relación donde Y tiene un conjunto de palabras Z, tal que ese conjunto Z se encuentra en el idioma X. Entonces se cumpliría que las palabras en ese conjunto, son comunes entre Y y X.

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar ambos idiomas en los campos de texto respectivos. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Operaciones idioma e idioma		
Idioma X	Idioma Y	Resultado
eng	afr	8

Cantidad palabras comunes

Listar palabras comunes

Idioma que más aporte a otro

Porcentajes de aporte

Note que en la salida se espera un valor numérico, indicando la cantidad de palabras en común, las cuales denominamos como el conjunto Z en el ejemplo anterior.

Operación 10

Descripción: Listar todas las palabras comunes entre dos idiomas

Lo que pretende esta funcionalidad es que a partir de dos idiomas (**X**, **Y**) proporcionados por el usuario, el sistema brinde una respuesta indicando la palabras que se encuentran relacionadas de manera común entre los lenguajes (**X** and **Y**).

Entonces, suponiendo que el usuario proporciona el lenguaje '**X = eng**' y '**Y = afr**,' el sistema verifica en la base de conocimiento para ver si existe un hecho que cumpla la relación donde **Y** tiene un conjunto de palabras **Z**, tal que ese conjunto **Z** se encuentra en el idioma **X**. Entonces se cumpliría que las palabras en ese conjunto, son comunes entre **Y** y **X**.

Utilización de la interfaz: Para llevar a cabo esta funcionalidad, basta con ingresar ambos idiomas en los campos de texto respectivos. Un ejemplo de este ingreso se encuentra en la siguiente imagen:

Operaciones idioma e idioma		
Idioma X	Idioma Y	Resultado
eng	afr	Palabras comunes
bok baasskap aardwolf Ciskei boomslang ag apartheid aardvark		

Cantidad palabras comunes

Listar palabras comunes

Idioma que más aportó a otro

Porcentajes de aporte

Note que en la salida se espera una lista de palabras. De no encontrar palabras, se despliega un mensaje indicando dicho resultado.

Operación 11

Descripción: Idioma que más aportó a otro.

Esta opción permite consultar la base de conocimiento para determinar cual lenguaje ha aportado el mayor porcentaje de palabras o derivaciones a otro lenguaje. El cálculo de dicho porcentaje se hace al contar la cantidad de palabras que un determinado lenguaje aportó al otro y posteriormente diviendo entre el total de palabras que ha recibido dicho lenguaje. Esto se realiza con todos los lenguajes existentes dentro de la base de conocimiento.

Utilización de la interfaz: Para utilizar esta funcionalidad no es necesario ingresar ningún lenguaje, al ejecutarse de esta manera, el porcentaje es calculado entre todos los lenguajes disponibles. Pero es posible definir el idioma **Y**, para así calcular el mayor porcentaje de aporte únicamente entre los aportes realizados al idioma **Y**.

Operaciones idioma e idioma

Idioma X	Idioma Y	Resultado	
		Máximo global	Cantidad palabras comunes
Idioma: ara a: xaa con			Listar palabras comunes
			Idioma que más aportó a otro
			Porcentajes de aporte

Operaciones idioma e idioma

Idioma X	Idioma Y	Resultado	
	ara	Máximo para: ara	Cantidad palabras comunes
a: ara a: ara con 64.238			Listar palabras comunes
			Idioma que más aportó a otro
			Porcentajes de aporte

Note que en la salida se espera un texto indicando el idioma origen del aporte, el idioma que recibió el aporte y el porcentaje que representa dicho aporte sobre el total de palabras recibidas. En caso de no haber aporte, la salida contendrá un texto indicando dicho resultado.

Operación 12

Descripción: Listar todos los idiomas que aportaron a otro.

Esta funcionalidad permite consultar todos los porcentajes de aporte que cada idioma a realizado a los demás, el porcentaje se calcula de la misma forma a como se describe en la operación anterior.

Utilización de la interfaz: Al igual que la operación anterior, no es necesario ingresar ningún lenguaje, lo que resulta en el cálculo de los porcentajes de todos los aportes realizados entre los idiomas. Pero también es posible definir el idioma Y, para así obtener los porcentajes de todos los aportes que cualquier idioma genera para dicho idioma Y.

Operaciones idioma e idioma

Idioma X	Idioma Y	Resultado	
		Todos los aportes	Cantidad palabras comunes
ang aporta a eng un 72			Listar palabras comunes
p_gmw aporta a ang un			Idioma que más aporte a otro
afr aporta a afr un 54.0'			Porcentajes de aporte
eng aporta a arz un 2.0			
ara aporta a fas un 100.			
ara aporta a ota un 100			
grc aporta a ara un 5.0'			
akk aporta a eng un 1.0			
ara aporta a gmh un 10			
aym aporta a arn un 12			

Operaciones idioma e idioma

Idioma X	Idioma Y	Resultado	
	eng	Aportes para: eng	Cantidad palabras comunes
afr aporta a eng un 3.0'			Listar palabras comunes
ara aporta a eng un 22.			Idioma que más aporte a otro
amh aporta a eng un 0			Porcentajes de aporte
ang aporta a eng un 72			
aii aporta a eng un 0.0'			
adt aporta a eng un 0.0			
arz aporta a eng un 0.0			
ale aporta a eng un 0.0			
abe aporta a eng un 0.0			
amj aporta a eng un 0.0			

Note que en la salida se esperan líneas de texto indicando los idiomas origen de cada aporte, con sus correspondientes idiomas receptores de cada aporte y el porcentaje que representa dicho aporte sobre el total de palabras recibidas para cada idioma. En caso de no haber aportes, la salida contendrá un texto indicando dicho resultado.

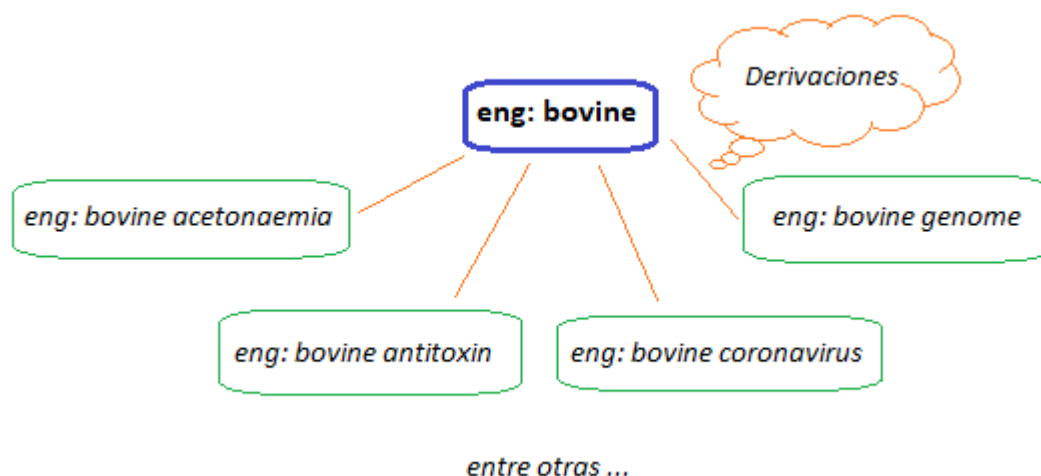
Resultados Interesantes

Operación 1

Descripción: Determinar si dos palabras son herman@s

En esta funcionalidad se determinó como dato curioso la palabra **eng: bovine**, de la cual derivan múltiples palabras, convirtiendo a estas en herman@s.

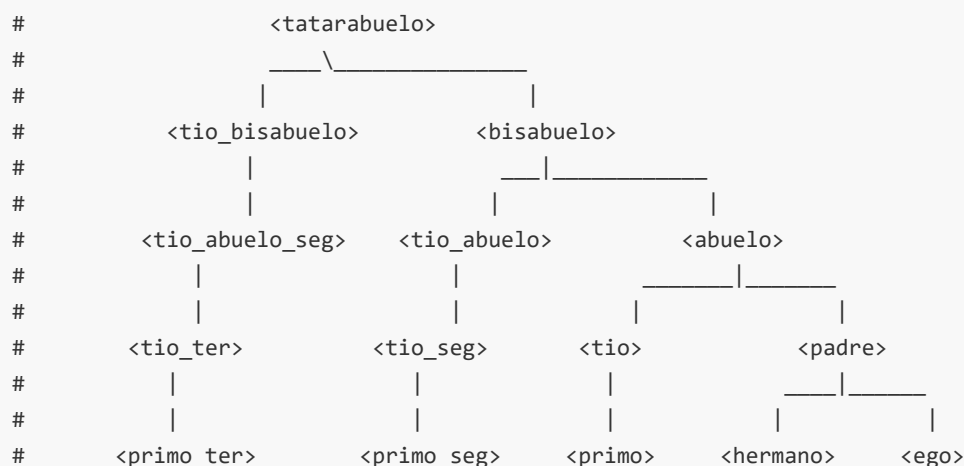
Un ejemplo claro de esta relación se muestra en la siguiente imagen:



Operación 2

Descripción: Determinar si dos palabras son prim@s

En principio, la implementación se realizó utilizando el tío directo de la palabra X, sin embargo se nota que la relación de primos se puede visualizar de una forma más general; como palabras que simplemente comparten un ancestro en común en una determinada cantidad de generaciones.

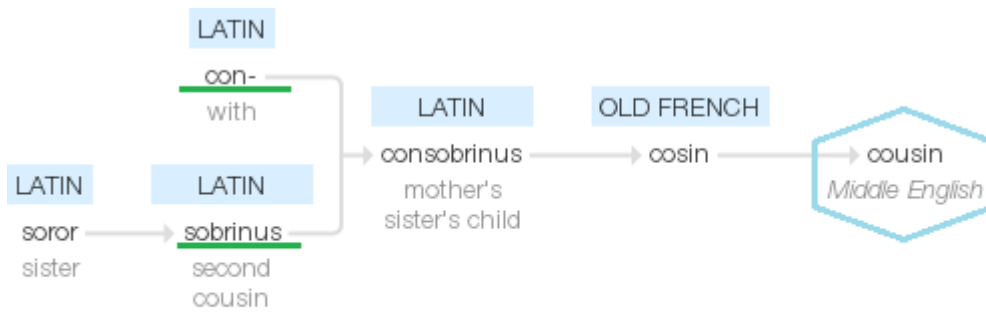


Operación 3

Descripción: Determinar si una palabra es hij@ de otra

Un dato curioso de esta funcionalidad lo encontramos con la palabra **consobrinus**, que es hij@ de: **con-** y **sobrinus**, que provienen del Latín. El dato más curioso en este caso, es que dicho hij@ produce la palabra **cousin** del inglés, la cula es sumamente reconocida y utilizada.

Para entender mejor este dato, basta con visualizar la siguiente imagen donde se muestran algunas de las relaciones anteriormente mencionadas.

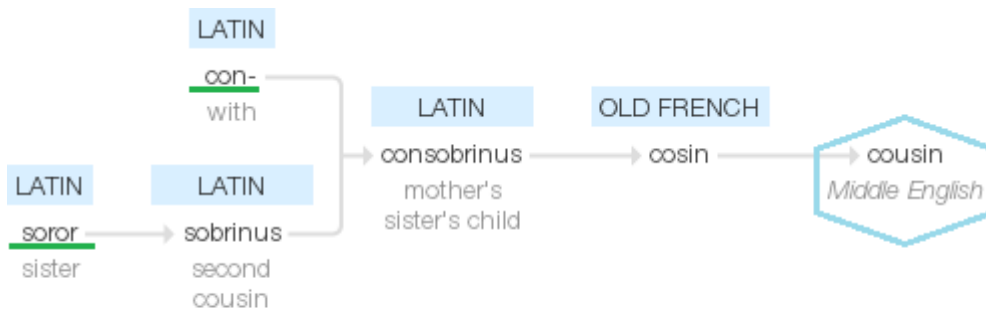


Operación 4

Descripción: Determinar si una palabra es ti@

El dato curioso para esta funcionalidad se puede basar en las relaciones de la *operación #3*, donde se encuentran las palabras: **con-** y **soror**, que provienen del Latín. El dato más curioso en este caso, es que *soror* es ti@ de *con* y de esta relación se produce la palabra **cousin** del inglés, la cula es sumamente reconocida y utilizada (tal como se mencionó en el caso anterior).

Para entender mejor este dato, basta con visualizar la siguiente imagen donde se muestran algunas de las relaciones anteriormente mencionadas.



Operación 5

Descripción: Determinar si son prim@s y en qué grado.

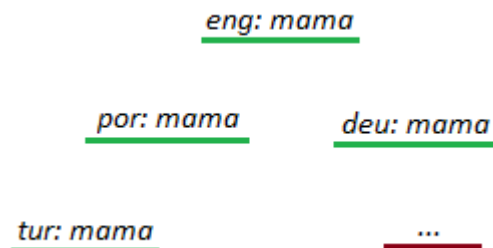
Para determinar el grado que tienen una pareja de primos es bastante sencillo si se entienden las relaciones tradicionales de la familia. Todos los primos se encuentran en un mismo nivel (visualizando el árbol de la operación 2). Una vez que se sabe que dos palabras (**X, Y**) son prim@, se puede determinar el grado (prim@os primeros, segund@s, tercer@s), contando cuantos ancestros tienen X y Y hasta llegar a un ancestro que tengan en común.

Operación 6

Descripción: Determinar si una palabra está relacionada con un idioma (Si / No)

Para esta funcionalidad específica el grupo de trabajo se dio a la tarea de buscar alguna palabra que se encontrara relacionada con múltiples lenguajes y que tuviera un significado similar en todos ellos. Y finalmente, se optó por **mama** que hace referencia a *mamá* en el lenguaje Español. Entonces, al ejecutar esta funcionalidad se encontraban bastantes casos donde dicha palabra está relacionada con un idioma determinado.

Ejemplos de estos idiomas se encuentra en la siguiente imagen:



Operación 7

Descripción: Obtener el conjunto de todas las palabras en un idioma originadas por una palabra específica

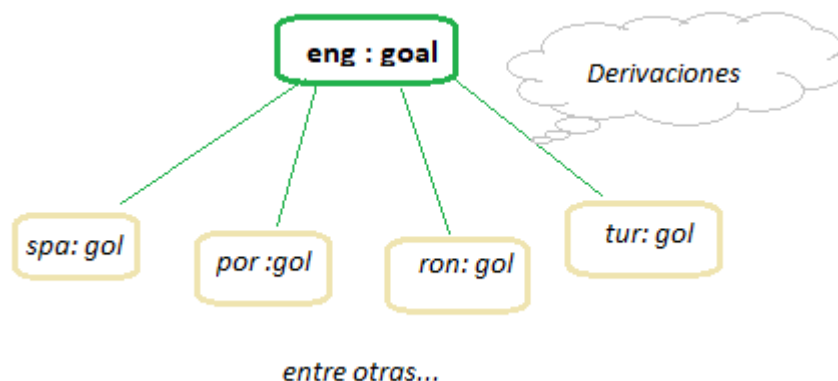
Un dato curioso en esta funcionalidad es que la palabra en latín **aequus**, que significa *igual, llano, justo, equilibrado, equitativo* tiene relación con las palabras en castellano *adecuado, ecuación, equilátero, equilibrio, entre otras*. Denotando así la importancia que tiene una simple palabra y lenguaje, para substraer palabras derivadas en un lenguaje destino específico.

Operación 8

Descripción: Listar los idiomas relacionados con una palabra

Una palabra que resulta ser muy interesante de analizar en este apartado es '**gol**', que está relacionada con '**eng : goal**' y que se relaciona con la temática del fútbol. ¿Pero porqué es interesante? Pues resulta que dicha palabra no varía entre múltiples lenguajes e inclusive tiene el mismo significado, por lo que brinda un valor agregado al utilizarla en esta funcionalidad.

A continuación, se muestra una imagen con algunos ejemplos de la palabra '**gol**' relacionada con múltiples lenguajes:



Operación 9

Descripción: Contar todas las palabras comunes entre dos idiomas

Se realizó un conteo entre el idioma Malayo (usualmente hablado en Malasia) representado como **zsm** y un lenguaje criollo basado en él, conocido como Malayo de Ambon o también Ambonés (hablado en la isla de Ambon, que se encuentra en un grupo de islas de Indonesia) y que se representa como **abs**. A pesar de que el segundo se basa en el primero, en la base de datos únicamente se encontró una relación según lo indicado por la operación 9.

Dado que es difícil de creer que dichos lenguajes tengan tan poco en común, esto debe significar simplemente que los datos no están nada completos. La incompletitud de los datos era esperada, pero lo interesante de todo esto es que en la base se haya incluido una única palabra de relación entre ambos idiomas tan cercanos.

Operación 10

Descripción: Listar todas las palabras comunes entre dos idiomas

Habiendo obtenido el resultado mencionado en la operación 9, se procedió a listar la palabra en común entre ambos idiomas Malayo **abs** y el Ambonés **zsm**, con lo cuál se obtiene el resultado de que la palabra en común entre ambos idiomas es **beta**, lo curioso de esto es que además de lo mencionado en la operación 9, la única palabra que se decidió incluir en la base de datos fue **beta**. Una palabra que ni siquiera parece tener un significado destacado en los ambos lenguajes, pues su traducción es la misma en el idioma Español. Entonces ¿por qué incluir una palabra que no tiene mayor importancia en la base de datos y no incluir ninguna otra palabra?

Operación 11

Descripción: Idioma que más aportó a otro.

El resultado más llamativo encontrado no es específico a un lenguaje o porcentaje. Parte de las instrucciones para el proyecto, era el medir el aporte realizado por lenguaje mediante el porcentaje de palabras que este aportó a otro. El problema de esta medición radica en las grandes diferencias entre lenguajes en la base de datos. Existen lenguajes que aparecen en más de medio millón de relaciones, mientras que otros aparecen únicamente en dos relaciones. Lo anterior significa que lenguajes que aportan muy pocas palabras pueden tener un 100% de aporte a otros lenguajes, mientras que otros que aportan cincuenta mil palabras pueden llegar únicamente a un 30% o 50% de aporte.

En resumen, la métrica de usar un porcentaje provee información distorsionada si no se está consciente de la realidad de los datos. En caso de que esta comparación fuera realmente importante, sería de más utilidad el conformar una métrica que también incluya la cantidad de palabras aportadas.

Operación 12

Descripción: Listar todos los idiomas que aportaron a otro.

Durante la definición de los hechos para las pruebas se identificó una situación interesante, en el caso de existir un lenguaje **A**, que aporta la palabra **ab** para el lenguaje **X**, y al mismo tiempo existir un lenguaje **B** que aporta la palabra **ab**, al mismo lenguaje **X**; si en los datos, para el lenguaje **X** únicamente existe la palabra **ab**, entonces surge la siguiente duda: ¿Aportaron ambos lenguajes **A** y **B** un 100% al lenguaje **X**, ambos aportaron **0%**, o aportó 100% el lenguaje **A** y el lenguaje **B** no aportó nada o viceversa?

Dada la situación se tomó la decisión de que ambos lenguajes aportan (un 100% en el caso del ejemplo hipotético) al lenguaje **X**. Esto pues no se encontró un criterio claro para definir a que lenguaje debe atribuirse el aporte, puesto que no hay mucha información a la mano. Además de que la decisión no tiene mayor importancia a simple vista.

Detalles de Implementación y Diseño

Diseño

Respecto al diseño del sistema, se decide crear una especie de módulos con una división basada en el modelo MVC (Model, View, Controller), no obstante, también se incluyen folders para la incorporación de las pruebas, los archivos utilizados y herramientas.

Entonces de esta manera, el módulo V (view) se encarga de controlar todo lo relacionado con el diseño, despliegue y ejecución de la interfaz gráfica mostrada al usuario. Que incluye eventos por medio de botones, inputs para ingresar datos y áreas destinadas a mostrar las respuestas recibidas por parte del sistema.

Por otra parte, el módulo C(Controller) se encarga del flujo de acceso entre la vista y el modelo, entonces lo que pretende es brindar la definición de funciones que reciben los parámetros respectivos para ejecutar alguna de las reglas definidas en la parte del modelo.

La sección del modelo, contiene todo lo relacionado a la lógica, es donde se definen las reglas y donde se hacen las inferencias respectivas a cada método.

El módulo de **test**, contiene todas las pruebas desarrollas por medio de pytest para cada una de las funcionalidades que contempla el proyecto.

Por último, la carpeta **util** almacena código que está relacionado a herramientas, entre estas se encuentran las de lectura y escritura de archivos, entre otras.

Implementación

En este apartado es importante mencionar que no todas las relaciones que se encontraban en el archivo etymwn.tsv que representa la base de datos, fueron consideradas, ya que el equipo de trabajo encontró que algunas relaciones eran recíprocas, entonces con solo una de ellas bastaría para encontrar la relación.

La lista de relaciones que se consideraron a lo largo del desarrollo de este proyecto son:

Relación	Registros
etymology	473 515
etymological_origin_of	473 433
etymologically_related	538 558
has_derived_form	2 264 744
Total:	3 750 250

Para dar un ejemplo claro de porqué se tomó esta decisión es la eliminación de la relación: **is_derived_from**, que es recíproca a la relación **has_derived_form**.

- Esto se concluyó por medio de la verificación a nivel de datos, de que realmente si existe la regla **has_derived_form**, también estaba la contraparte **is_derived_from**, con los valores de lenguaje y palabra correspondientes.
- Inclusive si lo vemos a nivel general, podemos determinar que existe la misma cantidad de relaciones **is_derived_from** que de **has_derived_form**, lo que también da un indicio fuerte de que los registros son recíprocos.

Operaciones 11 y 12

Para las operaciones 11 y 12, las instrucciones se interpretaron de dos formas diferentes, pues no habían indicios claros de cuál interpretación es la correcta. La primera interpretación señala que el cálculo de los porcentajes de tenía que hacerse para todos los lenguajes que aportaban a un lenguaje específico definido como parámetro. Es decir, dado un lenguaje específico, calcular los porcentajes de aporte de todos los demás lenguajes a este específico.

La segunda interpretación señala que el cálculo de los porcentajes tenía que hacerse en un cruce entre todos los lenguajes disponibles. Es decir, se identificaron 397 lenguajes y en caso de cada lenguaje haber aportado al menos una palabra a cada uno de los demás, se deberían calcular "factorial de 397" porcentajes de aporte.

Dado que no había claridad sobre cual debía implementarse, se implementaron ambas versiones.

Manejo de los tipos de relaciones

Parte de las instrucciones del proyecto incluye que en la interfaz gráfica sea posible seleccionar cuáles relaciones se consideran al ejecutar cada operación de consulta. Lo anterior debe poder realizarse sin reiniciar por completo la aplicación, es por eso que se definen hechos específicos que poseen un valor booleano dependiendo de que relación debe tomarse en cuenta para la consulta. Los hechos definidos son:

etymology_active(True / False)

has_derived_form_active(True / False)

etymologically_related_active(True / False)

etymological_origin_of_active(True / False)

Distribución de Trabajo

La distribución de trabajo en este proyecto se realizó según las funcionalidades que fueron solicitadas, es decir que a cada uno de los tres estudiantes le tocó desarrollar un conjunto de estas.

- Palabra con Palabra
- Palabra con lenguaje
- Lenguaje con lenguaje

Respecto a los integrantes del proyecto:

Nombre	Carné	Nota
Brandon Dinarte Chavarría	2015088894	100
Armando López Cordero	2015125414	100
Julian Salinas Rojas	2015114132	100

Estudiantes de Ingeniería en Computación del Instituto Tecnológico de Costa Rica.