



Grupo 7

Gabriel Amaral 4.590.503-3

Diego Melli 4.747.143-8

Ezequiel 4.899.156-9

## Contenido

[Descripción del problema](#)

[Objetivos](#)

[Descripción de la solución](#)

[Tecnologías utilizadas](#)

[Servidor y Aplicación Web Administrador:](#)

[Aplicación iOS:](#)

[Openshift:](#)

[Java-JBoss-JPA:](#)

[Implementación y distribución](#)

[BackOffice:](#)

[Capturas de Pantalla del BackOffice:](#)

[Front Office:](#)

[Capturas de Pantalla del Front Office móvil:](#)

[BackEnd:](#)

[Web Services](#)

[Crear Usuario](#)

[Nueva sentencia para el usuario](#)

[Siguiendo sentencia para el usuario](#)

[Formato Json](#)

[Sentence](#)

[WordsDef](#)

[Word](#)

[User](#)

[jsonNextSentence](#)

[Desafíos](#)

[Modelo](#)

[Obtener Material](#)

[Alternativas encontradas:](#)

[Priorización de los textos](#)

[Pensando a futuro](#)

[Conclusiones](#)

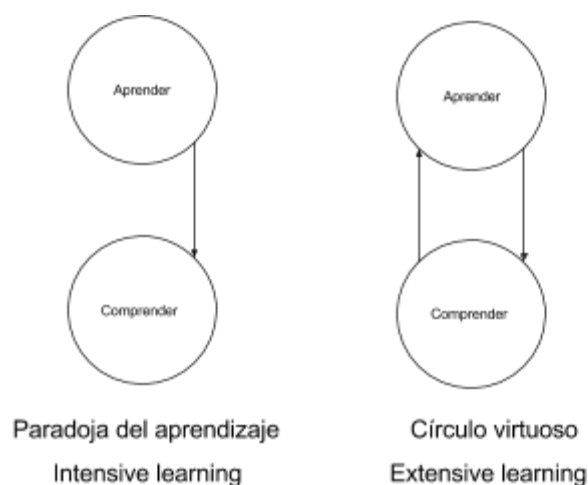
[Referencias](#)

## Descripción del problema

Aprender un nuevo idioma siempre es un gran reto, conlleva mucho esfuerzo y dedicación por un largo período de tiempo (en general varios años). Esta dificultad se acentúa cuando el lenguaje a aprender es muy diferente a nuestro lenguaje nativo, imposibilitando la reutilización de los conocimientos en otros lenguajes que ya conocemos.

A lo largo de este proceso de aprendizaje, encontrar material que se adecúe a nuestro nivel actual es uno de los grandes desafíos. Cuando el tiempo pasa y no vemos resultados la motivación baja y es posible que el estudiante deje de estudiar el idioma.

Una posible explicación a este fenómeno está en la paradoja del aprendizaje. En este modelo se diferencian dos importantes instancias en el aprendizaje de un idioma, una es de aprendizaje y otra de comprensión. La primera se refiere a el aprendizaje propio del lenguaje. La segunda etapa es de comprensión, donde se aplica lo aprendido en la inmersión en el lenguaje, utilizando el lenguaje de manera práctica en actividades tales como conversaciones, lectura de libros, noticias, páginas web, escritura, etc. Obviamente esta etapa de comprensión genera nuevo aprendizaje naturalmente. El problema que se encuentra es que, hasta alcanzar un cierto nivel en el lenguaje, es muy difícil lograr la comprensión y utilización ya que no disponemos aún del vocabulario y dominio de la gramática para hacerlo. Esto deriva en que al principio, el aprendizaje sea artificial, basado en libros, pocos ejemplos de alta dificultad utilizando lo que aprendimos. A este método de aprendizaje se le denomina *intensive learning* y es el que generalmente se utiliza en las clases y en los programas de aprendizaje de idiomas.



Luego de llegado a un cierto nivel del lenguaje (la literatura lo sitúa en aproximadamente dos mil palabras), se da la comprensión lográndose una inmersión en el lenguaje y pudiendo utilizarlo en la vida diaria. A esta etapa del aprendizaje se le denomina Círculo Virtuoso en el que el aprendizaje se da naturalmente. Contrariamente a cuando nos encontramos en la paradoja del aprendizaje, a esta forma de aprendizaje se le denomina

*extensive learning*, en el que se entrena con una gran cantidad de ejemplos, elegidos por el aprendiz, de baja dificultad (ej: conversaciones con pares, películas, noticias, etc).

El problema que buscamos atacar es esta misma paradoja, que dificulta en gran medida el progreso del aprendizaje.

## Objetivos

Los objetivos que nos propusimos son los siguientes:

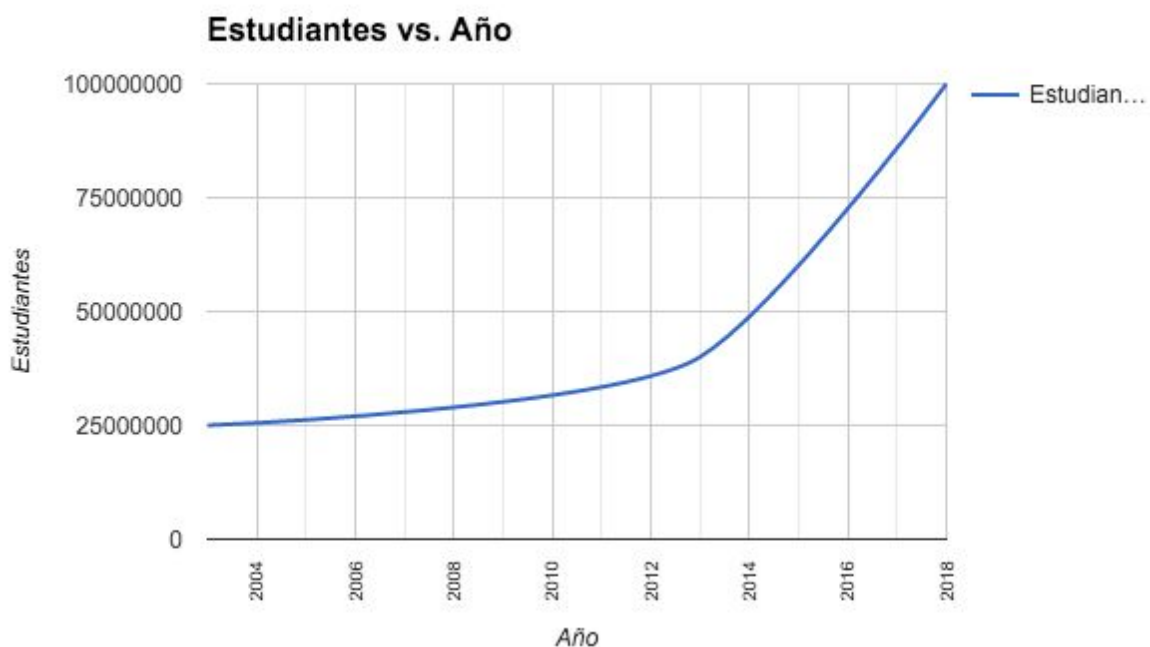
- Mejorar la capacidad de comprensión de textos
- Aumentar la velocidad de lectura
- Aumentar vocabulario

Plasmando esto en que, luego de practicar todos los días por un tiempo, el usuario estará en condiciones de poder:

- Navegar en internet en chino
- Leer noticias en chino
- Poder moverse sin problemas en China

## Descripción de la solución

Nos enfocamos en atacar los efectos de la paradoja del aprendizaje. Para acotar la problemática, en una primera instancia, decidimos enfocar los esfuerzos en un único lenguaje de aprendizaje. El idioma elegido fue el chino, debido a su alta popularidad y su altas proyecciones de crecimiento. Elegimos inglés como idioma de la app, por lo tanto se asume que el usuario maneja el inglés con fluidez.



*Evolución de personas estudiando chino y proyecciones al 2018*

También nos focalizamos únicamente en la lectura como herramienta de aprendizaje, en donde el conocimiento es fácilmente extensible a otras áreas del lenguaje.

El objetivo es permitir que el usuario logre entrar en menos tiempo en el círculo virtuoso antes descrito, permitiéndole poder leer noticias, navegar en internet, comunicarse por chat, poder moverse sin problemas en China, y así, aprender de forma natural.

Para poder lograr estos objetivos planteados nos basamos en los siguientes métodos con eficacia comprobada:

- [\*Extensive reading\*](#). Premisa: lectura de abundantes textos que el lector pueda comprender mejoran sustancialmente la velocidad de lectura y la comprensión de textos (sobre el cual se han escrito varios papers científicos).
- [\*Spaced repetition\*](#). Premisa: la retención de palabras es más efectivo cuanto su práctica se distribuye de forma espaciada en el tiempo.

Se desarrolló un sistema que se adapta al nivel del usuario y velocidad de aprendizaje, en base a estructuras de datos y métodos de aprendizaje automáticos que se adaptan de acuerdo al constante y "seamless" feedback del usuario utilizando heurísticas.

La metodología es sencilla: presentar textos de fácil comprensión para el usuario dado su nivel actual y que ésta maximicen el aprendizaje del lector.

De esta forma, y en base al uso diario, se espera potenciar la habilidad de comprender textos (y de forma más rápida).

Idealmente:

Valor: Usuario x Texto → Real

- que sea entendible por el usuario
- que maximice el aprendizaje de nuevas palabras

De todas los textos, se devuelve el que más valor de aprendizaje posea.

La ventaja de este sistema es se libera de la responsabilidad de elegir lo próximo que leer al usuario, presentando lo que más optimice su aprendizaje y que el usuario podrá percibir como su capacidad de comprensión de textos y vocabulario aumenta con el tiempo (junto a la extensión y dificultad de los textos que le aparecen para leer).

Se plantean dos modelos para la aplicación :

- **Exploración:** Se explora el actual nivel del usuario presentando en su mayoría palabras nuevas
- **Explotación:** Se prioriza mostrar frases con palabras sabidas para maximizar la probabilidad de comprensión

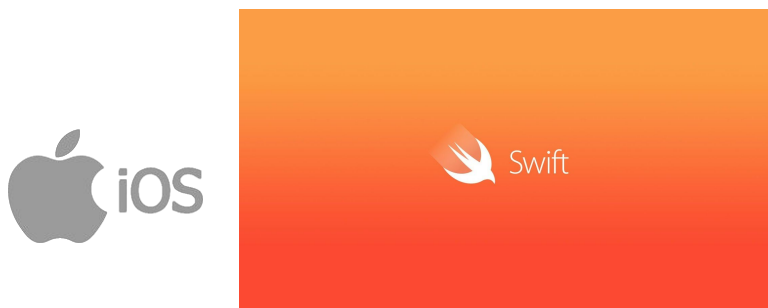
## Tecnologías utilizadas

- Servidor y Aplicación Web Administrador:





- Aplicación iOS:



Descripción general:

Openshift:

Es una plataforma orientada a servicios (PaaS) que permite desarrollar rápidamente, aplicaciones escalables en un entorno de nube.

Fue utilizado para alojar el en la nube el proyecto java que contiene el backoffice web y los web Services rest para ser consumidos por el móvil.

Java-JBoss-JPA:

**Java EE:** es una plataforma de programación para desarrollar y ejecutar software de aplicaciones en el lenguaje de programación Java.

**JBoss:** es un servidor de aplicaciones Java EE de código abierto implementado en Java puro.

**JPA:** es una interfaz java que describe el manejo de los datos relacionales.

Este conjunto de tecnologías se utilizaron para la realización del backend. Se programó en lenguaje java y se usó jboss como servidor de aplicaciones.

Jpa fue usado para manejar el mapeo de objetos a una forma relacional para ser persistidos en una base de datos relacional, en este caso mySql.

## Implementación y distribución

El sistema cuenta con 3 módulos, backoffice(administración), frontoffice(usuario) y BackEnd(implementación de las funcionalidades).

## BackOffice:

Consta de una página web que permite a los administradores el manejo de información referente a usuarios, sentencias y words.

Capturas de Pantalla del BackOffice:

id	Mail
1	
2	admin@admin.com
3	
4	
5	
6	
348	h@h.com

Listado de Usuarios del sistema

id	Level	Translation
12	VERY_BASIC	Thank you. You are the best!
23	VERY_BASIC	I'm not from here.
28	VERY_BASIC	He is very tall.
35	VERY_BASIC	Taiwan is a very safe place.
42	VERY_BASIC	I'm mad at you.
54	VERY_BASIC	Love all, play!
60	VERY_BASIC	You decide!
69	VERY_BASIC	Please say it one more time.
75	VERY_BASIC	What are you busy with?
86	VERY_BASIC	Which part of China are you from?
90	VERY_BASIC	No way!, Really?! (expressing disbelief)
94	VERY_BASIC	Where are you from?
102	VERY_BASIC	Time passes so quickly!
109	VERY_BASIC	I'm really full!
118	VERY_BASIC	What else do we need to buy?
126	VERY_BASIC	What is the difference?
135	VERY_BASIC	Why are you so late?
144	VERY_BASIC	What makes you think so?
155	VERY_BASIC	Now she lives in the U.K.

Listado de sentencias



Bu (步)		admin@admin.com		log out
Know Words				
Add Sentences				
All Words				
All Sents				
All Users				

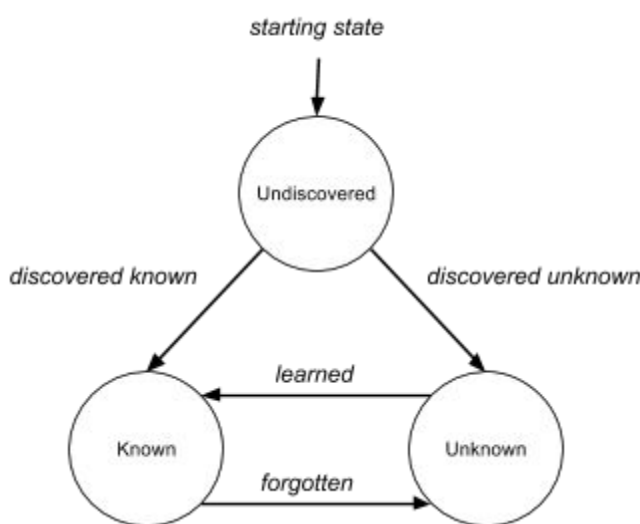
  

Words				
Hanzi	Pinyin	Link		difficulty
谢	xiè	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=7736">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=7736</a>		BASIC
谢谢	xiè xiè	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=34">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=34</a>		VERY_BASIC
你	nǐ	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=12">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=12</a>		VERY_BASIC
谢谢你	xiè xiè nǐ	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=1531">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=1531</a>		BASIC
最	zuì	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=318">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=318</a>		VERY_BASIC
好	hǎo	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=183">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=183</a>		VERY_BASIC
最好	zuì hǎo	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=562">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=562</a>		BASIC
了	le	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=53">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=53</a>		VERY_BASIC
我	wǒ	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=1">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=1</a>		VERY_BASIC
不	bù	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=15285">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=15285</a>		VERY_BASIC
是	shì	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=4">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=4</a>		VERY_BASIC
不是	bú shì	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=305">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=305</a>		VERY_BASIC
这	zhè	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=745">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=745</a>		VERY_BASIC
裡	lǐ	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=695">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=695</a>		VERY_BASIC
這裡	zhè lǐ	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=77">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=77</a>		VERY_BASIC
的	de	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=62">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=62</a>		VERY_BASIC
人	rén	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=25">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=25</a>		VERY_BASIC
他	tā	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=30">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=30</a>		VERY_BASIC
很	hěn	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=10">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=10</a>		VERY_BASIC
高	gāo	<a href="https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=1787">https://www.learnwitholiver.com/chinese/show-word-details.php?word_id=1787</a>		VERY_BASIC

Listado de Words.

### Front Office:

Es una aplicación móvil (inicialmente para iPhone), en forma de flashcards y con un ambiente que favorezca la comprensión del texto: con definición de las palabras y traducción del material de modo de maximizar la probabilidad de comprensión de la idea. Al presionar el usuario una palabra, aparecerá su definición y pronunciación. Está quedará marcada como no sabida (utilizando la heurística razonable que el usuario solo presiona las palabras que no sabe para obtener la definición). De esta forma, al cambiar el texto, se memorizan las palabras que el usuario sabe y las palabras que el usuario no sabe.



En todo momento se mantendrá y actualizará los conjuntos de palabras: Undiscovered, Unknown, Known. De la movilidad de estos se podrán obtener estadísticas de progreso que

les serán mostradas al usuario y comparadas con sus objetivos personales (por ejemplo, rendir un examen oficial, o entender el 20% de las palabras más comunes).

A su vez, para cada frase se cuenta con la posibilidad de mostrar la traducción y marcarla como favorita. El objetivo de la primera es despejar dudas cuando un texto es complejo, concluyendo que el texto presentado al usuario es más complejo de lo que debe. La segunda es para tener una idea de los gustos del usuario, intentando traer frases que sean similares a las frases favoritas

Capturas de Pantalla del Front Office móvil:





### BackEnd:

Para la comunicación entre el móvil (Front Office) y el BackEnd se definieron servicios Rest, que intercambian información en formato json.

### Web Services

Funcion	Crear Usuario
Url	<a href="http://rest-buu.rhcloud.com/services/getNewUser">http://rest-buu.rhcloud.com/services/getNewUser</a>
Tipo	get
Parámetros	no
Retorno	Json: User { "id":13, "mail":null, "known":null, "unknown":null }

Observaciones	13 es un ejemplo, retorna la id autogenerada por la base de datos

Funcion	Nueva sentencia para el usuario
Formato	getNewSentence/{idU}
Tipo	Get
Url	<a href="http://rest-buu.rhcloud.com/services/getNewSentence/5">http://rest-buu.rhcloud.com/services/getNewSentence/5</a>
Parámetros	idUsuario
Retorno	Json: Sentence
Observaciones	5 es un ejemplo de id de usuario

Funcion	Siguiente sentencia para el usuario
Formato	nextSentence
Tipo	Post
Url	<a href="http://rest-buu.rhcloud.com/services/nextSentence/">http://rest-buu.rhcloud.com/services/nextSentence/</a>
Parámetros	no
Retorno	Json: Sentence
Cuerpo	Json: jsonNextSentence

## Formato Json

### Sentence

```
{  
  "id": Long,  
  "translation": String,  
  "level": Enum,  
  "link": String,  
  "wordsDef": WordsDef[]  
}
```

### WordsDef

```
{  
  "id": Long,  
  "here": String,  
  "Word": Word,  
}
```

### Word

```
{  
  "id": Long,  
  "hanzi": String,  
  "pinyin": String,  
  "link": String,  
  "difficulty": Enum,  
  "definition": String[],  
  hasWords: wordsDef[],  
}
```

### User

```
{  
  "id": Long,  
  "mail": String!,  
  "known": Word[],  
  "unknown": Word[]  
}
```

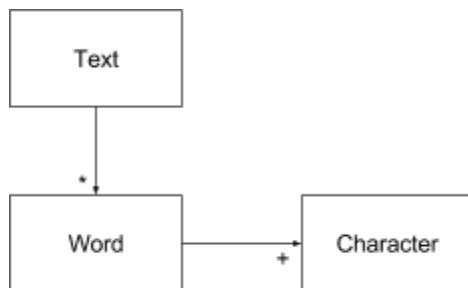
```
jsonNextSentence
```

```
{  
  user: Long, //id del usuario  
  sentence: Long, //id de la sentencia  
  unknowWords: word[] //Long[] id de las words no conocidas  
}
```

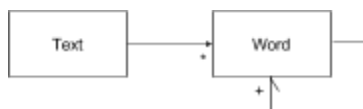
## Desafíos

### Modelo

Nos costó plasmar en un modelo las relaciones entre los textos, las palabras y los caracteres que capture la estructura del idioma. No deseábamos un modelo que sea excesivamente simplista que no nos permitiera extender luego el sistema. El primer modelo que pensamos y con el que empezamos a desarrollar fué el siguiente:



Luego, nos dimos cuenta que los caracteres podían ser palabras, debiendo manejar mismos objetos conceptuales en dos objetos de programación diferentes. Además perdíamos información debido a que no representamos que una palabra puede estar formado por más de una palabra. Resolvimos cambiar el modelo al siguiente:



Esto nos permite poder extender el sistema en un futuro y no duplicar entidades.

## Obtener Material

Un recurso fundamental para el buen funcionamiento de la aplicación es contar una gran base de texto traducido y etiquetados con sus definiciones.

### Alternativas encontradas:

- Generar nosotros mismos el material (Descartamos por falta de tiempo y conocimiento).
- Tercerizar la generación del contenido: Empleado, Crowdsourcing (Descartado por falta de tiempo y presupuesto).
- Partnerships con páginas que ofrezcan material etiquetado como páginas de aprendizaje de Chino. (Detalles a continuación)
- Scrapping (sin permiso) de una página de aprendizaje con mucho material etiquetado (learn witholiver.com). (Illegal) ← Fue el elegido para el prototipo
- Generación de la base de forma automática (Descartado por su complejidad, detalles a continuación)

Luego, debido a que el material está bajo Copyrights, nos comunicamos con el CEO de Antosch & Lin Languages, la organización por detrás de la herramienta LearnWithOliver de aprendizaje de idiomas, para intentar llegar a un acuerdo para poder utilizar comercialmente su base de textos a cambio de regalías. Luego de intercambiar varios emails prefirió no acceder, pero pidió seguir en contacto. No intentamos buscar otro socio ya que decidimos explorar la opción de generar la base de textos automáticamente.

Encontramos una plataforma colectiva de traducciones entre diversos idiomas llamado Tatoeba. Los textos traducidos del Chino a Inglés (por hablantes nativos) son aproximadamente 40 mil. La licencia (CC By) permite su uso para fines comerciales solo citando al autor.

Un ejemplo de estos textos:

我還年輕的時候曾經祈禱，希望自己能得到一輛自行車。後來我意識到，這樣向上帝祈求是不通的。因此我偷了一輛自行車，並祈求上帝能寬恕我。

*When I was young I used to pray for a bike. Then I realized that God doesn't work that way. So I stole a bike and prayed for forgiveness.*

Como se puede ver, es necesario tokenizar el texto en chino y luego tener un diccionario para cada una de esas palabras. Encontramos varias alternativas gratis y eficaces para tokenizar el texto. Es necesario un diccionario Chino a Inglés. Encontramos una buena alternativa pero esto, un diccionario colaborativo de Chino a Inglés. El problema que encontramos es que cuenta con licencias Creative Commons By ShareAlike, que a pesar de permitir utilizarlo embebido en aplicaciones con fines comerciales, implica licenciar el sistema CC By ShareAlike. Actualmente estamos estudiando este tema comunicándonos con



el responsable de esta iniciativa y con abogados especialistas en el tema (profesores de la materia Aspectos éticos y legales del manejo de datos).

## Priorización de los textos

Como planteamos al principio, lo ideal sería tener una función:

Valor: Usuario x Texto → Real

- que sea entendible por el usuario
- que maximice el aprendizaje de nuevas palabras

Intentamos encararlo de esta forma (con una única función) pero luego concluimos que es demasiado complejo y no es eficiente del punto de vista computacional.

Finalmente nos resolvimos diseñando dos capas de decisión:

### Filtrado

Filtrado: Integer, Texto → Booleano

Filtrado(threshold, sentence) = true if #UnknownWords ≤ threshold

Factible(threshold, Sentencias) = {sentence | T(threshold, sentence) }

En otras palabras, el conjunto factible serán todas las frases que contengan menos de *threshold* palabras no conocidas por el usuario, limitando la dificultad de la misma.

### Priorización

Sobre este conjunto reducido de frases se aplica la función priorización

Priorización: Texto → Real

Esta priorización depende de:

- Largo de frase
- Las palabras desconocidas pero que el usuario conoce las palabras que la componen aumentan la priorización
- Repetir palabras que ya se repitieron
- Priorizar palabras que se saben pero no aparecieron por un largo tiempo

Elegimos dentro de las factibles la que tenga más valor.

En cuanto a la implementación del sistema, implementamos la función Filtrado, realizando directamente una consulta SQL para que devuelva los ids de los textos que cumplen con la condición, devolviendo una de los textos factibles. Nos resta implementar la función priorización que quedará para futuras iteraciones.

## A futuro

- Realizar un producto mínimo viable a partir de este prototipo y validarlo en la Appstore.
- Tener en cuenta los intereses del usuario para mostrarle sentencias relacionadas a temas de su interés.
- Generar una página que muestre el progreso de una forma clara y atractiva.
- Permitirle al usuario fijar sus metas personales (como pruebas de Chino internacionales, conocer el 20% de las palabras más usadas, leer determinado libro o artículo, etc) y medir el progreso dinámica micamente contra estas metas
- Agregar el aprendizaje automático para que el sistema verdaderamente se adapte a los diferentes requerimientos del usuario y no solo se adapte a el vocabulario conocido.
- Extenderlo a otros idiomas.
- Crear versiones para otras plataformas, por ejemplo android o windows phone.

## Conclusiones

Realizando este proyecto aprendimos mucho sobre varios temas entre los que se destacan:

Desarrollo de aplicación móvil en iOS

Desarrollo de backend y manejo intensivo de bases de datos

Negociación con posibles partners

Temas de propiedad intelectual y licencias

Logramos acercarnos lo más posible al alcance planeado, siendo muy ambicioso en un primer momento. Se implementó la aplicación llamada Bu, manteniendo las palabras conocidas y no conocidas para cada usuario y presentarle (o recomendarle) el texto a leer adaptándose de de forma dinámica al progreso del usuario. Además se intentó hacer la interfaz lo más amigable y con un interfaz de usuario atractiva, consiguiendo lo que se quería.

Seguiremos trabajando en el sistema hasta lograr la validación en el mercado.

## Referencias

- [www.learnwitholiver.com](http://www.learnwitholiver.com)
- <https://www.openshift.com>
- <https://www.mysql.com/>
- <http://getbootstrap.com/>
- <http://www.oracle.com/technetwork/java/javaee/tech/persistence-jsp-140049.html>
- [www.jboss.org](http://www.jboss.org)