

SheetChat: Generación de chatbots a partir de hojas de cálculo

Haritz Medina Camacho

Director: Óscar Díaz García

Máster en Sistemas Informáticos Avanzados

Agenda

- ▶ Contexto y problema
- ▶ Chatbot
- ▶ SheetChat
- ▶ Casos de estudio
- ▶ Conclusiones

Contexto y problema

Contexto y problema

Smartphone

- ▶ El Smartphone **ofrece movilidad, pero se reduce su usabilidad** debido a:
 - ▶ El contexto móvil hace complejo su uso: calle, ruido
 - ▶ Conectividad limitada y poco fiable
 - ▶ Tamaño de pantalla reducido
 - ▶ Limitación de procesamiento
 - ▶ Entrada de datos mediante: teclado táctil, gestos o voz
- ▶ Hay tareas muy demandadas difíciles de llevar a cabo, como **consultar hojas de cálculo**.

Contexto y problema

Hojas de cálculo

- ▶ Un estudio [Derek Flood, 2011] revela que **el 79% de personas** seleccionadas a través del European Spreadsheet Risk Interest Group **requieren de consultas a hojas de cálculo en una configuración móvil:**
 - ▶ Para consultas del día a día o cotidianas
 - ▶ Demostración de datos a clientes
 - ▶ Hojas de cálculo compartidas con compañeros de trabajo

Contexto y problema

Hojas de cálculo

- El problema: son difíciles de consultar las hojas de cálculo:
 - No se pueden **representar todos los datos** en una pantalla
 - Hay que hacer **scrolling vertical y horizontal** para consultar datos

NotasAsignaturas

View only

NotasAsignatura

View only

NotasAsigna

View only

NotasAsignaturas

View only

	A	B		D		G		J	K	L	M	N	O		
1	Alumno	EjerEcuac	1	EjerFunciones1	EjerF	1	erIntegrales	Ex	1	enDerivInt	NotaContinua	ContAprobado	ExamenFinal	NotaFinal	Aprobado
2	Adrian Arana		2	5.5	2		6.5	2		8	7.375	si	0	7.375	si
3	Santiago Rocha		3	6.5	3		7.5	3		3.5	3.85	no	6.1	6.1	si
4	Alberto Ballester		4	3.5	4		3.5	4		3	3.45	no	4	4	no
5	Jesus Barral		5	3.5	5		2.5	5		8	6.925	si	0	6.925	si
6	Alejandro Baquero		6	7	6		8	6		9	8.975	si	9.5	9.5	si
7	Maria Cuevas		7	7.5	7		9.5	7		4.5	4.7	no	4.5	4.7	no
8	Aurora Sabater		8	3	8		6.5	8		6	6.75	si	0	6.75	si
9	Esther Espina		9	6.5	9		7	9		5	4.9	no	6	6	si
10	Jaime Brito		10	4.5	10		3.5	10							

NotasMate

NotasFi

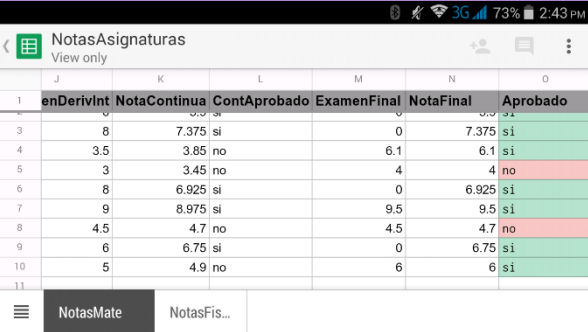
NotasMate

NotasMate

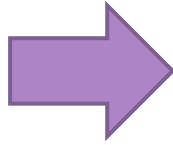
NotasFis...

Contexto y problema

SheetChat



	J	K	L	M	N	O
1	enDerivInt	NotaContinua	ContAprobado	ExamenFinal	NotaFinal	Aprobado
2	0	3.5	si	0	3.5	si
3	8	7.375	si	0	7.375	si
4	3.5	3.85	no	6.1	6.1	si
5	3	3.45	no	4	4	no
6	8	6.925	si	0	6.925	si
7	9	8.975	si	9.5	9.5	si
8	4.5	4.7	no	4.5	4.7	no
9	6	6.75	si	0	6.75	si
10	5	4.9	no	6	6	si

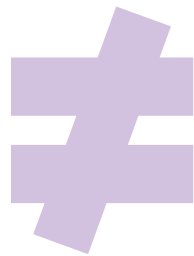
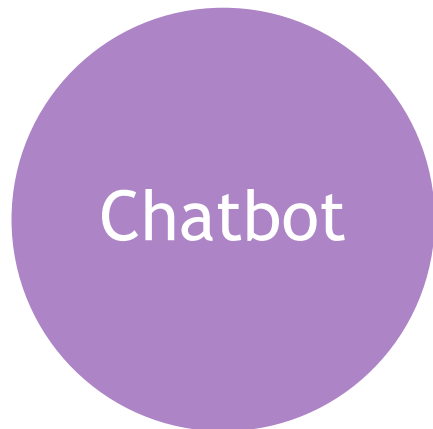


Chatbot

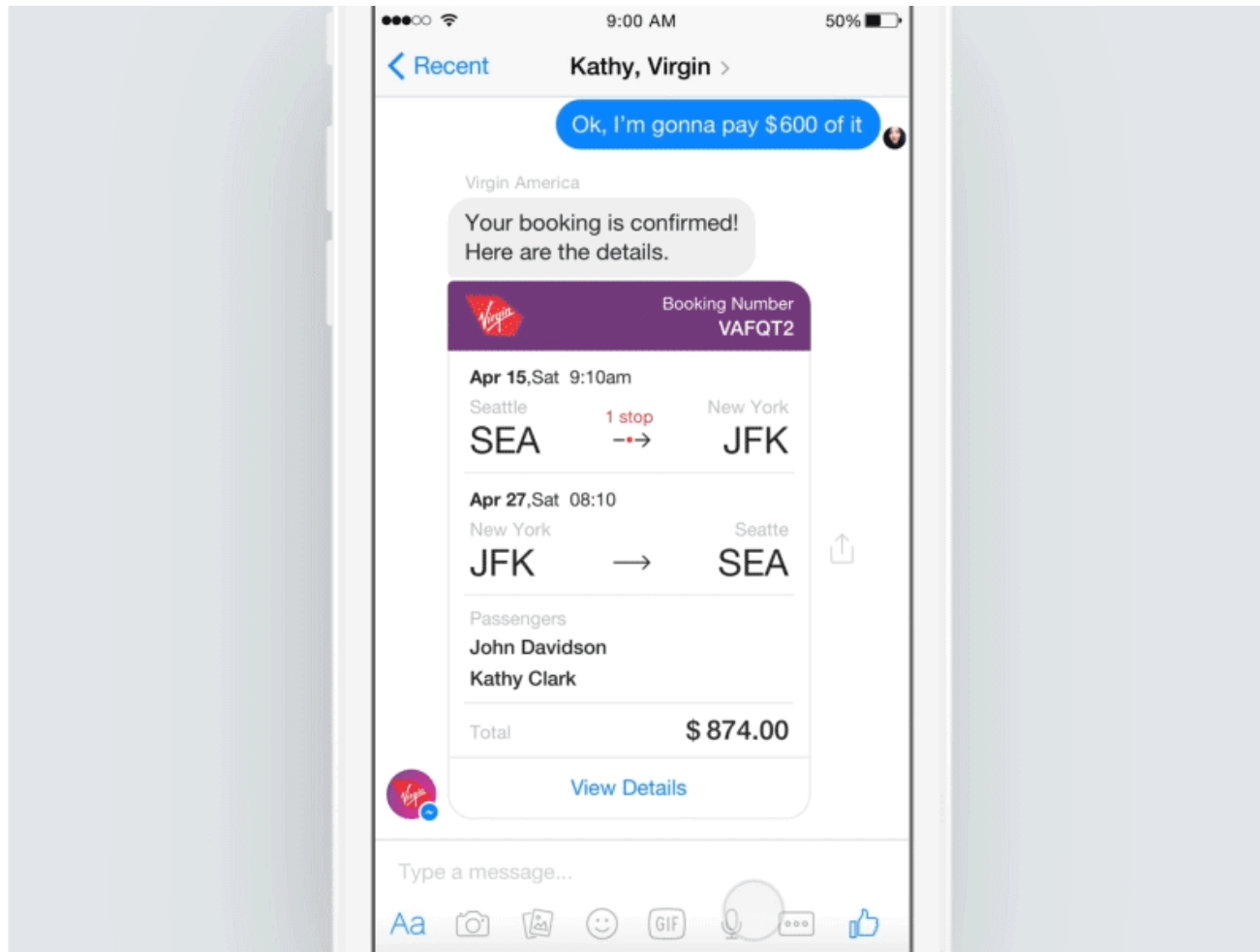
Chatbot

Definición

- ▶ Chatbot o bot conversacional o agente conversacional.
- ▶ Es un software que trata de simular una conversación con un humano con tal de automatizar tareas que actualmente realizan los usuarios por si mismos:
 - ▶ Reservar un restaurante
 - ▶ Añadir un evento al calendario
 - ▶ Buscar información



Chatbot Interacción



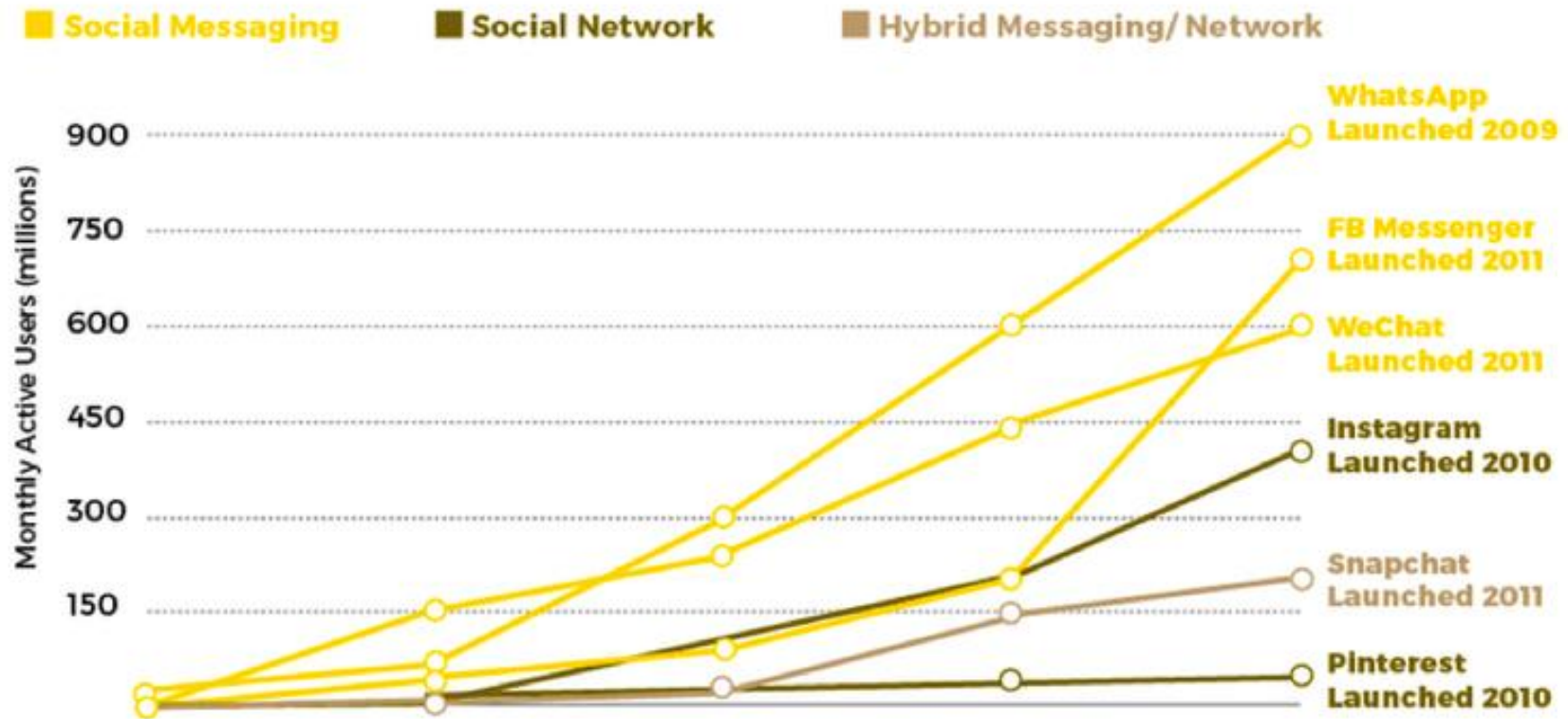
Chatbot

¿Por qué chatbots?

- ▶ El chat es una interfaz sencilla ya que los seres humanos interactuamos mediante conversaciones:
 - ▶ Es mucho más intuitivo que clicar en botones con un mouse
- ▶ Los jóvenes, que coparán el mercado en el futuro, dedican más tiempo al chat que a las redes sociales

Chatbot

¿Por qué chatbots?



Chatbot

Plataformas

- ▶ Plataforma conversacional: el servicio donde el usuario chatea con el chatbot.



- ▶ Plataforma de desarrollo: el servicio o librería utilizada para crear el chatbot
 - ▶ Telegram Bot API
 - ▶ Microsoft Bot Framework
 - ▶ Botkit

SheetChat

SheetChat

Objetivo

- ▶ Qué: creación de chatbots
- ▶ Quién: dueños de hojas de cálculo
- ▶ Para qué: facilitar la interfaz de consulta de **sus** hojas de cálculo
- ▶ Cómo: mediante un DSL (Lenguaje Específico de Dominio)

SheetChat

¿Qué es un DSL?



- ▶ Un DSL **es un lenguaje** de programación o especificación
- ▶ Está dedicado a resolver un problema concreto o a **un dominio concreto**
- ▶ Utiliza una notación y abstracciones concretas que **permiten una expresividad** más clara y fuerte
- ▶ En contraposición con los Lenguajes de Propósito General

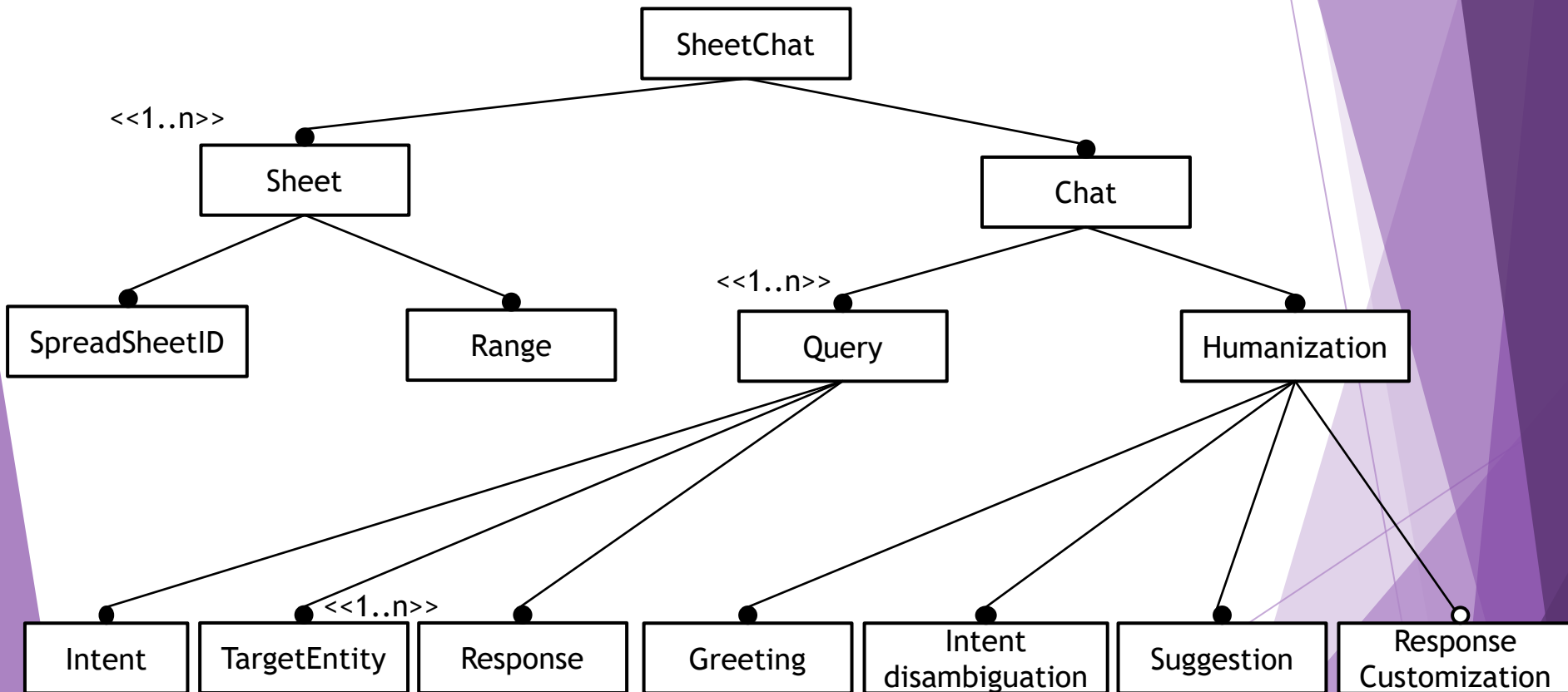
SheetChat

Research question

- ¿Cuáles son las abstracciones necesarias para captar la esencia de un chatbot que se utilice para consultar hojas de cálculo?

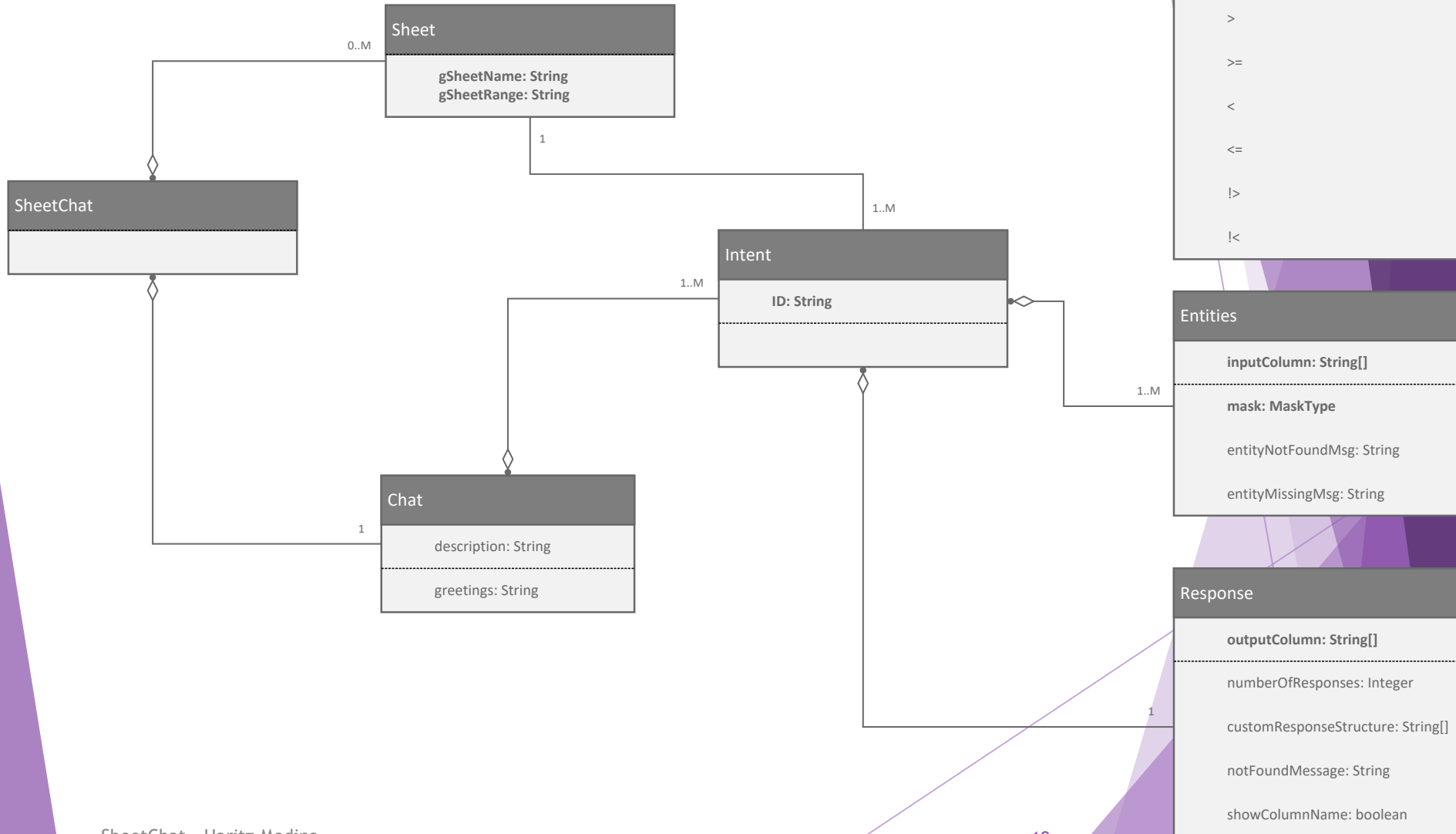
SheetChat

Modelo de características



SheetChat

Metamodelo



Casos de estudio

Casos de estudio

@notasbot

- ▶ Un profesor dispone de las notas de sus alumnos almacenadas en hojas de cálculo: matemática y física
- ▶ Utiliza la evaluación continua:
 - ▶ Trabajos y exámenes parciales durante el curso
 - ▶ Un examen final para los suspensos en continua
- ▶ Las preguntas habituales que realiza sobre las notas:
 - ▶ ¿Qué nota ha sacado Adrian en un parcial concreto?
 - ▶ ¿Qué nota ha sacado Aurora en los exámenes?
 - ▶ ¿Qué nota media ponderada lleva Esther en los trabajos?
 - ▶ ¿Ha aprobado Alberto la evaluación continua de Matemática?

Casos de estudio

@notasbot - Hoja de cálculo

NotasAsignaturas ☆

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

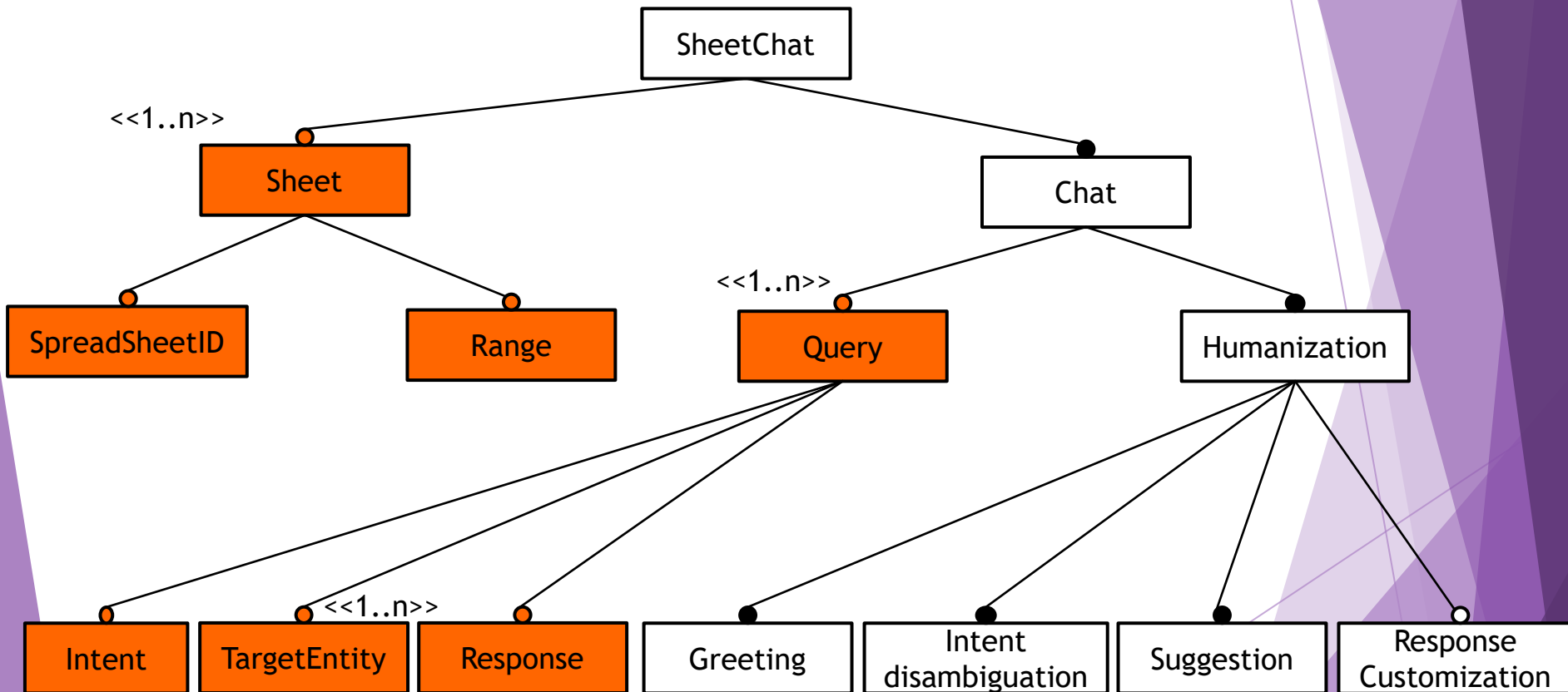
fx

	A	B	C	D	E	F	G	H	I	J	K	L
1	Alumno	EjerEcuaciones	EjerTrigonometria	EjerFunciones1	EjerFunciones2	EjerDerivadas	EjerIntegrales	ExamenEcuaciones	ExamenFunciones	ExamenDerivInt	NotaContinua	ContAprobado
2	Adrian Arana	6	7	5.5	6	7	6.5	4.5	5.5	6	5.9	si
3	Santiago Rocha	8	6	6.5	6.5	7	7.5	8.5	7	8	7.375	si
4	Alberto Ballester	4.5	4	3.5	3	2.5	3.5	4	5	3.5	3.85	no
5	Jesus Barral	5.5	4	3.5	3	1.5	2.5	6	2.75	3	3.45	no
6	Alejandro Baquero	5	5.5	7	7.5	7	8	6	7	8	6.925	si
7	Maria Cuevas	8	9	7.5	9.5	9	9.5	8.5	9.5	9	8.975	si
8	Aurora Sabater	6	4.5	3	4.5	5.5	6.5	4.5	4	4.5	4.7	no
9	Esther Espina	8	7.5	6.5	4.5	6.5	7	8	6.5	6	6.75	si
10	Jaime Brito	5	5.5	4.5	4	4	3.5	5.5	5.5	5	4.9	no
11												
12												
13												

+ ≡ | NotasMate ▾ NotasFisica ▾

Casos de estudio

@notasbot - Consultas en el DSL



¿Ha aprobado Alberto la evaluación continua de Matemática?

A	B	C	D	J	K	L
Alumno	EjerEcuaciones	EjerTrigonometria	EjerFuncion	ExamenDerivInt	NotaContinua	ContAprobado
Adrian Arana	6	7		6	5.9	si
Santiago Rocha	8	6		8	7.375	si
Alberto Ballester	4.5	4		3.5	3.85	no
Jesus Barral	5.5	4		3	3.45	no
Alejandro Baquero	5	5.5		8	6.925	si
Maria Cuevas	8	9		9	8.975	si
Aurora Sabater	6	4.5		4.5	4.7	no
Esther Espina	8	7.5		6	6.75	si
Jaime Brito	5	5.5		5	4.9	no

SELECT ContAprobado, Alumno
FROM NotasMate
WHERE Alumno LIKE "%alberto%"

```
{
  "ID": "mateContinuaAprobadoPorAlumno",
  "sourceSheet": "NotasMate",
  "response": {
    "outputColumn": "Alumno, ContAprobado",
    "notFoundMessage": "No se han encontrado resultados para el alumno.",
    "numberOfResponses": 1,
    "customResponseStructure": ["El alumno %s %s ha aprobado la evaluación"],
  },
  "entities": [{
    "inputColumn": "Alumno",
    "mask": "LIKE",
    "entityMissingMessage": "¿Que alumno quieres saber si ha aprobado la o",
    "entityNotFoundMessage": "Lo siento, no he encontrado informacion sobr",
  }],
}
```



```
"sheets": [
  {
    "gSheetName": "NotasMate",
    "gSheetRange": "A1:O10"
  },
  {
    "gSheetName": "NotasFisica",
    "gSheetRange": "A1:L10"
  }
],
"chat":{
  "description": "Este bot ayuda a obtener las notas de un alumno.",
  "greetings": "Hola! Soy notasbot y te ayudaré a obtener tus calificaciones. Prueba preguntandome por tus no
  "intents": [
    {
      "ID": "mateContinuaAprobadoPorAlumno",
      "sourceSheet": "NotasMate",
      "response":{
        "outputColumn": "Alumno, ContAprobado",
        "notFoundMessage": "No se han encontrado resultados para el alumno.",
        "numberOfResponses": 1,
        "customResponseStructure": ["El alumno %s %s ha aprobado la evaluación continua de matemática.", "Alu
      },
      "entities": [{
        "inputColumn": "Alumno",
        "mask": "LIKE",
        "entityMissingMessage": "¿Que alumno quieres saber si ha aprobado la continua de mate? Pregunta por e
        "entityNotFoundMessage": "Lo siento, no he encontrado informacion sobre ese alumno. Pregunta por ejem
      }]
    },
  ],
}
```

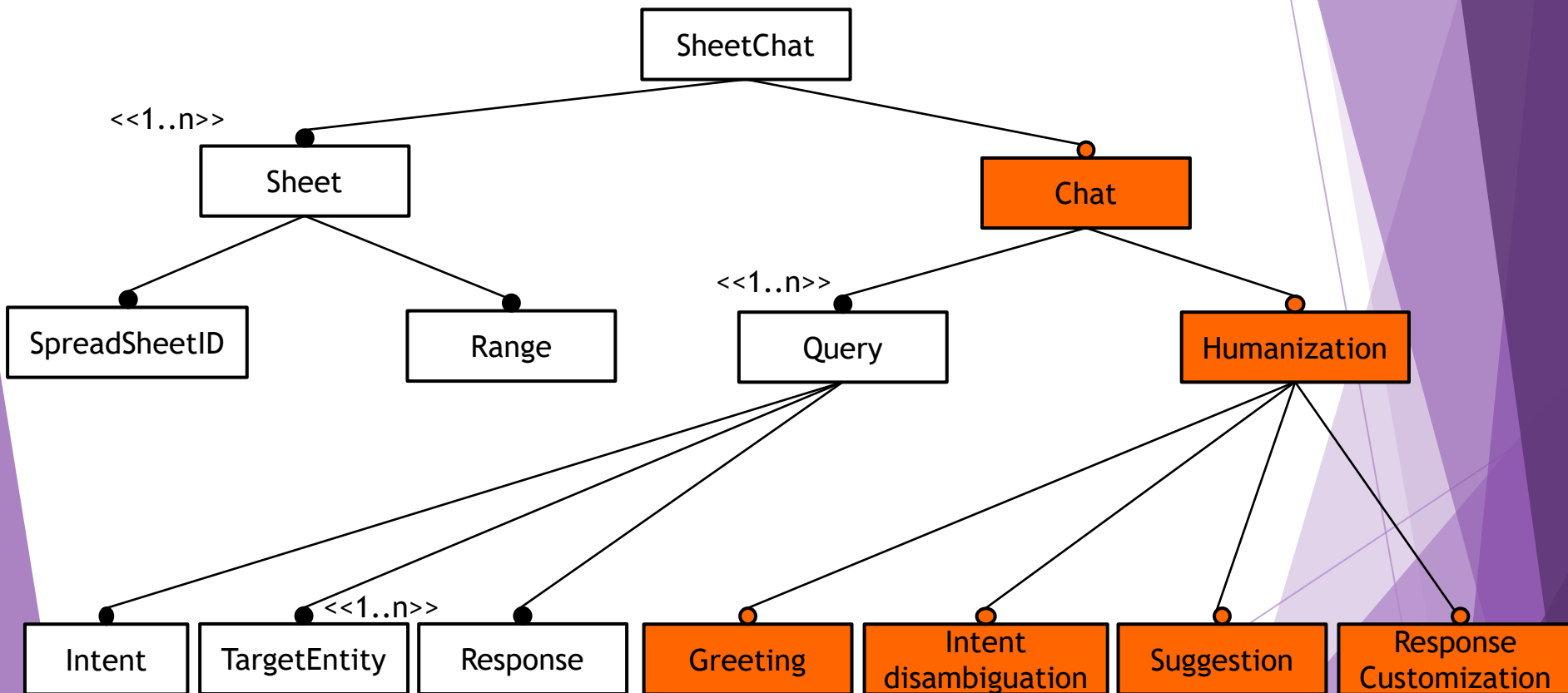
Intent

Respuesta

Entidades

Casos de estudio

@notasbot - Humanización en el DSL



```
"sheets": [
  {
    "gSheetName": "NotasMate",
    "gSheetRange": "A1:O10"
  },
  {
    "gSheetName": "NotasFisica",
    "gSheetRange": "A1:L10"
  }
],
```

```
"chat":{
  "description": "Este bot ayuda a obtener las notas de un alumno.",
  "greetings": "Hola! Soy notasbot y te ayudaré a obtener tus calificaciones. Prueba preguntandome por tus no
```

Saludo

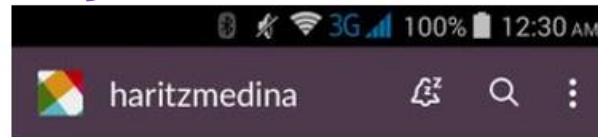
```
"intents": [
  {
    "ID": "mateContinuaAprobadoPorAlumno",
    "sourceSheet": "NotasMate",
    "response":{
      "outputColumn": "Alumno, ContAprobado",
      "notFoundMessage": "No se han encontrado resultados para el alumno.",
      "numberOfResponses": 1,
      "customResponseStructure": ["El alumno %s %s ha aprobado la evaluación continua de matemática.", "Alu
```

Respuesta personalizada

```
},
"entities":[{
  "inputColumn": "Alumno",
  "mask": "LIKE",
  "entityMissingMessage": "¿Que alumno quieres saber si ha aprobado la continua de mate? Pregunta por e
  "entityNotFoundMessage": "Lo siento, no he encontrado informacion sobre ese alumno. Pregunta por ejem
}]
},
```

Casos de estudio

@notasbot - Ejecución de la consulta



@notasbot



haritzmedina 12:21 AM

Hola



notasbot 12:21 AM

Hola! Soy notasbot y te ayudaré a obtener tus calificaciones. Prueba preguntandome por tus notas de matematica o fisica



haritzmedina 12:22 AM

Ha aprobado alberto la evaluacion continua de mate?



notasbot 12:22 AM

El alumno Alberto Ballester no ha aprobado la evaluación continua de matemática.



haritzmedina 12:22 AM

Que nota tiene en los exámenes de mate alberto?



notasbot 12:22 AM



Message @notasbot



Saludo

Consulta

Casos de estudio

@notasbot - Ejecución de la consulta



Sugerencias

Casos de estudio

@conferencebot

- ▶ Un asistente al congreso WISE 2015 dispone de un calendario con las sesiones que desea consultar durante el congreso
- ▶ Algunas preguntas habituales durante el congreso son:
 - ▶ ¿Qué charlas hay hoy a las 15:00?
 - ▶ ¿Cuándo son las charlas sobre seguridad web?
 - ▶ ¿En qué sesiones participa una persona?

Casos de estudio

@conferencebot - Hoja de cálculo

	A	B	C	D	E	F
1	Day	Slot	Session1	Session2	Session3	Session4
2	Sunday, November 1st	7:30-9:00	Room: Paris Breakfast			
3	Sunday, November 1st	9:00-10:00	Room: Paris Chair: S.S. Iyengar Keynote 1: Hardware Accelerators for the Web			
4	Sunday, November 1st	10:30-12:00	Room: Niece Chair: Jie Shao S1: Big Data Techniques	Room: Cannes Chair: Devis Bianchini S2: Social issues of Web Applications	Room: Marseille Chair: Yanchun Zhang S3: Deep/Hidden Web	Room: Paris Chair: Deren Chen & William Wei Song SS1: Data Quality and Trust in Big Data
5	Sunday, November 1st	12:00-13:30	Room: Paris Lunch			
6	Sunday, November 1st	13:30-15:00	Room: Niece Chair: Dingding Wang S4: Big Data and Mining Techniques	Room: Cannes Chair: Marek Rusinkiewicz S5: Social Network Computing	Room: Marseille Chair: Eduardo B. Fernandez cure Web Systems Architectures Using Se	Room: Paris Chair: Shaozhong Zhang SS2: Data Quality and Trust in Big Data
7	Sunday, November 1st	15:30-17:00	Room: Niece Chair: Hideyuki Takada S6: Big Data Applications	Room: Cannes Chair: Frederique Laforest S7: Social Web	Room: Marseille Chair: Eduardo B. Fernandez cure Web Systems Architectures Using Se	Room: Paris Chair: Siril Yella SS3: Data Quality and Trust in Big Data
8	Sunday, November 1st	17:30-18:30	Room: Paris Reception			
9	Monday, November 2nd	7:30-9:00	Room: Paris Breakfast			
10	Monday, November 2nd	9:00-10:00	Room: Monte Carlo Chair: Tao Li science Research at FIU (S. S. Iyengar, Naphtali Rishe, Steve Luis, and Radu Jianu)			
11	Monday, November 2nd	10:30-12:00	Room: Niece Chair: Zhiguo Gong S8: Web Data Integration and Mashups	Room: Cannes Chair: Shu-Ching Chen S9: Web Privacy and Security	Room: Monte Carlo Chair: Richi Nayak S10: Integration of Web with IoT	
12	Monday, November 2nd	12:00-13:30	Room: Paris Lunch			
13	Monday, November 2nd	13:30-15:00	Room: Niece Chair: Junhu Wang S11: Web Data Models	Room: Cannes Chair: Dimitrios Theodoratos S12: Web Information Retrieval	Room: Monte Carlo Chair: Byeong Ho Kang S13: Web-based recommendations	
14	Monday, November 2nd	15:00-17:00	Room: Niece Chair: Toshiyuki Amagasa	Room: Cannes Chair: Wojciech Cellary	Room: Monte Carlo Chair: Mark A. Finlayson	

Casos de estudio

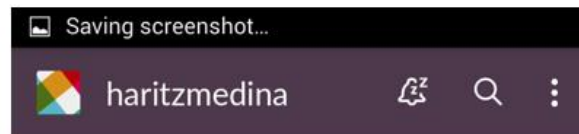
@conferencebot - Definición DSL

```
],  
"chat":{  
  "description": "WISE conference sessions related bot.",  
  "greetings": "This bot helps you to retrieve information related with the WISE 2015 c  
  "intents": [  
    {  
      "ID": "retrieveSlotByTopic",  
      "sourceSheet": "Schedules",  
      "response": {  
        "outputColumn": "Day, Slot",  
        "notFoundMessage": "Events not found",  
        "numberOfResponses": 5,  
        "customResponseStructure": ["An event at %s on %s is scheduled.", "Slot", "Day"  
      },  
      "entities": [  
        {  
          "inputColumn": ["Session1", "Session2", "Session3", "Session4"],  
          "mask": "LIKE",  
          "entityMissingMessage": "Search for a topic, room or chairman, for example:",  
          "entityNotFoundMessage": "Topic, room or chairman not found, try another one:  
        }  
      ]  
    },  
  ],  
}
```

Búsqueda en múltiples columnas

Casos de estudio

@conferencebot - Ejecución de la consulta



@conferencebot

Session3: Room: Marseille
Chair: Yanchun Zhang
S3:Deep/Hidden Web
Session4: Room: Paris
Chair: Deren Chen
& William Wei Song
SS1:Data Quality and Trust in Big Data



haritzmedina 3:04 AM
when are scheduled web sessions?



conferencebot 3:04 AM
An event at 9:00-10:00 on Sunday,
November 1st is scheduled.

An event at 10:30-12:00 on Sunday,
November 1st is scheduled.

An event at 13:30-15:00 on Sunday,
November 1st is scheduled.

An event at 15:30-17:00 on Sunday,
November 1st is scheduled.

An event at 10:30-12:00 on Monday,
November 2nd is scheduled.

Consulta



Message @conferencebot



Casos de estudio

@conferencebot

Demo

Casos de estudio

@tripadvisor

- ▶ Consulta de restaurantes en Miami con Tripadvisor.
- ▶ Se utiliza import.io para extraer hojas de cálculo de una búsqueda de Tripadvisor.
- ▶ Preguntas habituales a la hora de buscar restaurantes:
 - ▶ ¿Puedes mostrarme restaurantes con comida vegana?
 - ▶ ¿Puedes darme información sobre restaurantes italianos más baratos de 15\$?

Casos de estudio

@tripadvisor - Hoja de cálculo

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	url	Image	Image_alt	Image_link	Name	Name_link	Ranking	NumberOfReview	NumberOfReview	Review1	Review1_link	Review1Date	Review2
2													

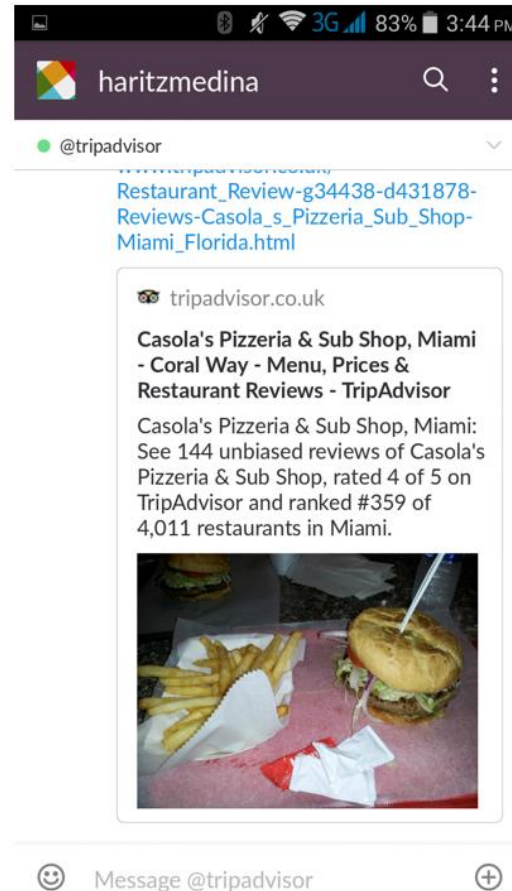
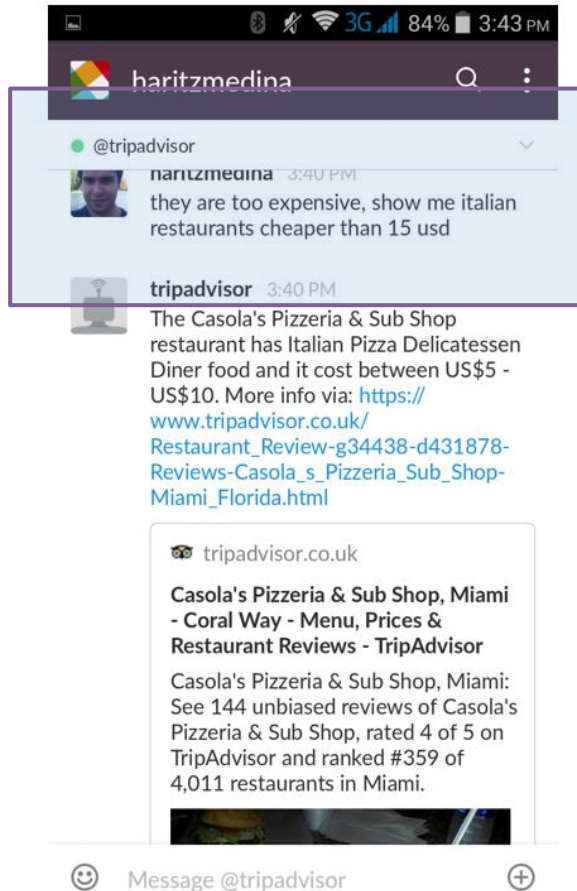
Número de resultados a mostrar

```
{
  "ID": "restaurantsCheaperThanPerCuisine",
  "sourceSheet": "Restaurants",
  "response": {
    "outputColumn": "Name, RangePrice, Cuisines, Name_link",
    "numberOfResponses": 3,
    "customResponseStructure": ["The %s restaurant has %s food and it cost between %s.",
      "Name", "Cuisines", "RangePrice", "Name_link"]
  },
  "entities": [
    {
      "inputColumn": "Cuisines",
      "mask": "LIKE",
      "entityMissingMessage": "Which type of cuisine would you like? For example:",
      "entityNotFoundMessage": "Cuisine style not found, try with:"
    },
    {
      "inputColumn": "MaxPrice",
      "mask": "<",
      "entityMissingMessage": "Set maximum price. For example:",
      "entityNotFoundMessage": "Not found cheaper than that price. Can you set a higher"
    }
  ]
}
```

Comparación de valores numéricos

Casos de estudio

@tripadvisor



Casos de estudio

@tripadvisor

Demo

Conclusiones y trabajo futuro

Trabajo futuro

- ▶ Evaluación de la herramienta
- ▶ Generación del DSL mediante una interfaz gráfica o web
- ▶ Proporcionar más recursos para mejorar la usabilidad de los chatbots generados
- ▶ Estudio del ecosistema de plataformas

Conclusiones

- ▶ Definir patrones o buenas prácticas para la generación de chatbots ayuda a reducir el coste de aprendizaje de uso del bot
- ▶ Herramientas NLP y AI actuales como Wit.ai permiten desarrollar chatbots capaces de interpretar lenguaje natural con poco entrenamiento

Bibliografía

- ▶ Flood, D., Harrison, R., & Daid, K. M. (2011). Spreadsheets on the Move : An Evaluation of Mobile Spreadsheets.
- ▶ Flood, D. (2010). Natural language processing for spreadsheet information retrieval. *Natural Language Processing and Information Systems*, 5723, 309-312.
http://doi.org/10.1007/978-3-642-12550-8_32
- ▶ Montag, C., Błaszkiwicz, K., Sariyska, R., Lachmann, B., Andone, I., Trendafilov, B., ... Markowetz, A. (2015). Smartphone usage in the 21st century: who is active on WhatsApp? *BMC Research Notes*, 8, 331.
<http://doi.org/10.1186/s13104-015-1280-z>

