75.74

Sistema Distribuidos 1

Paulo César Cuneo

84840

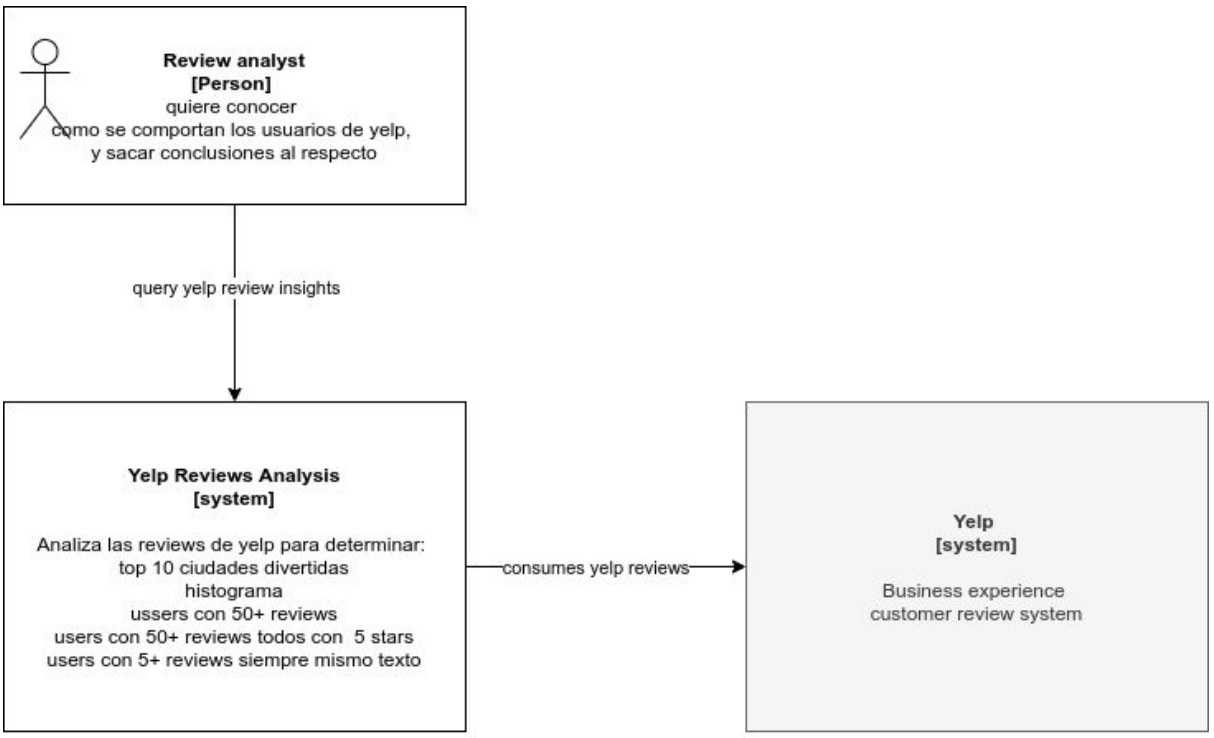
TP2

Yelp Reviews Analysis

Correcciones

Índice

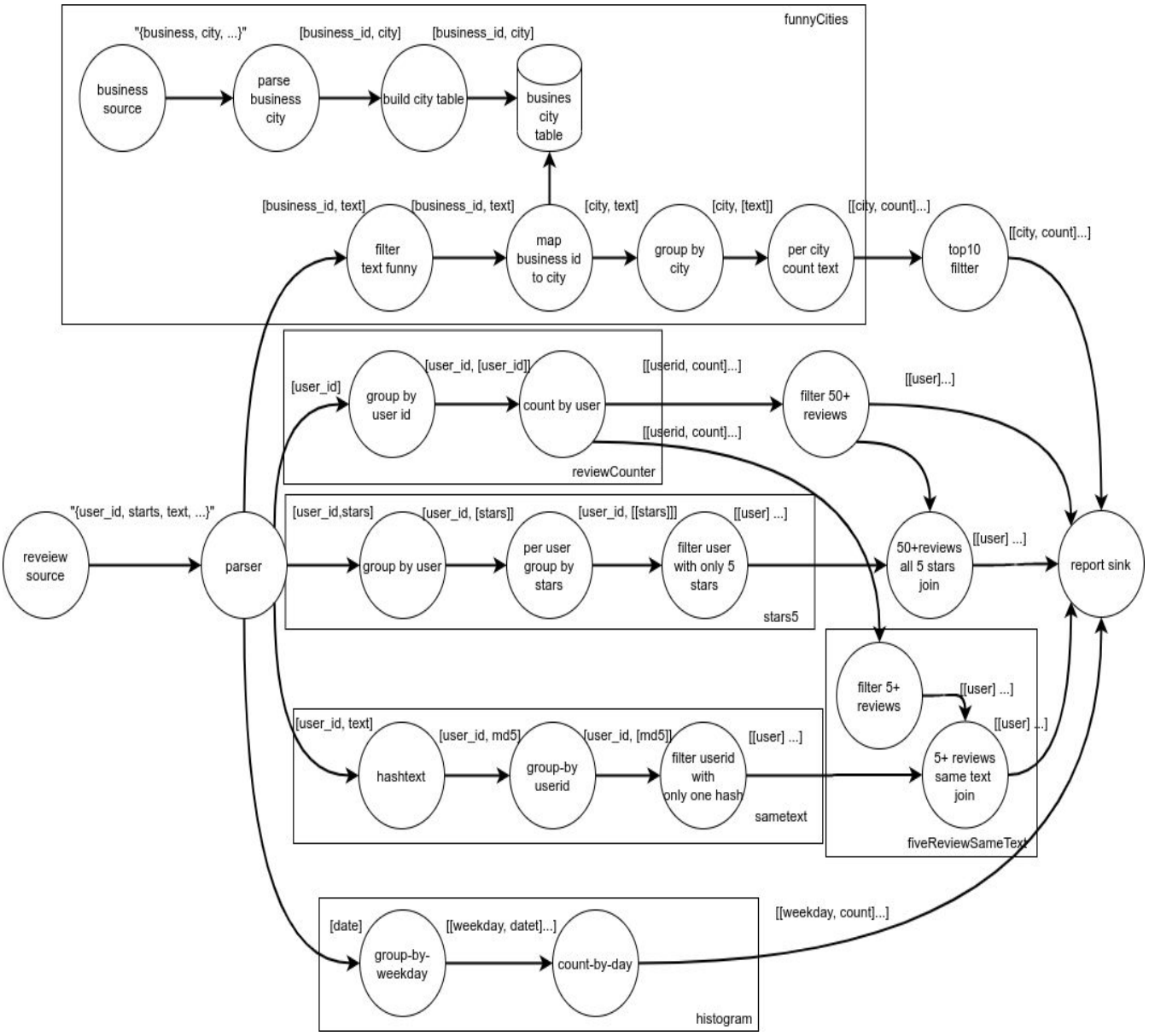
Context



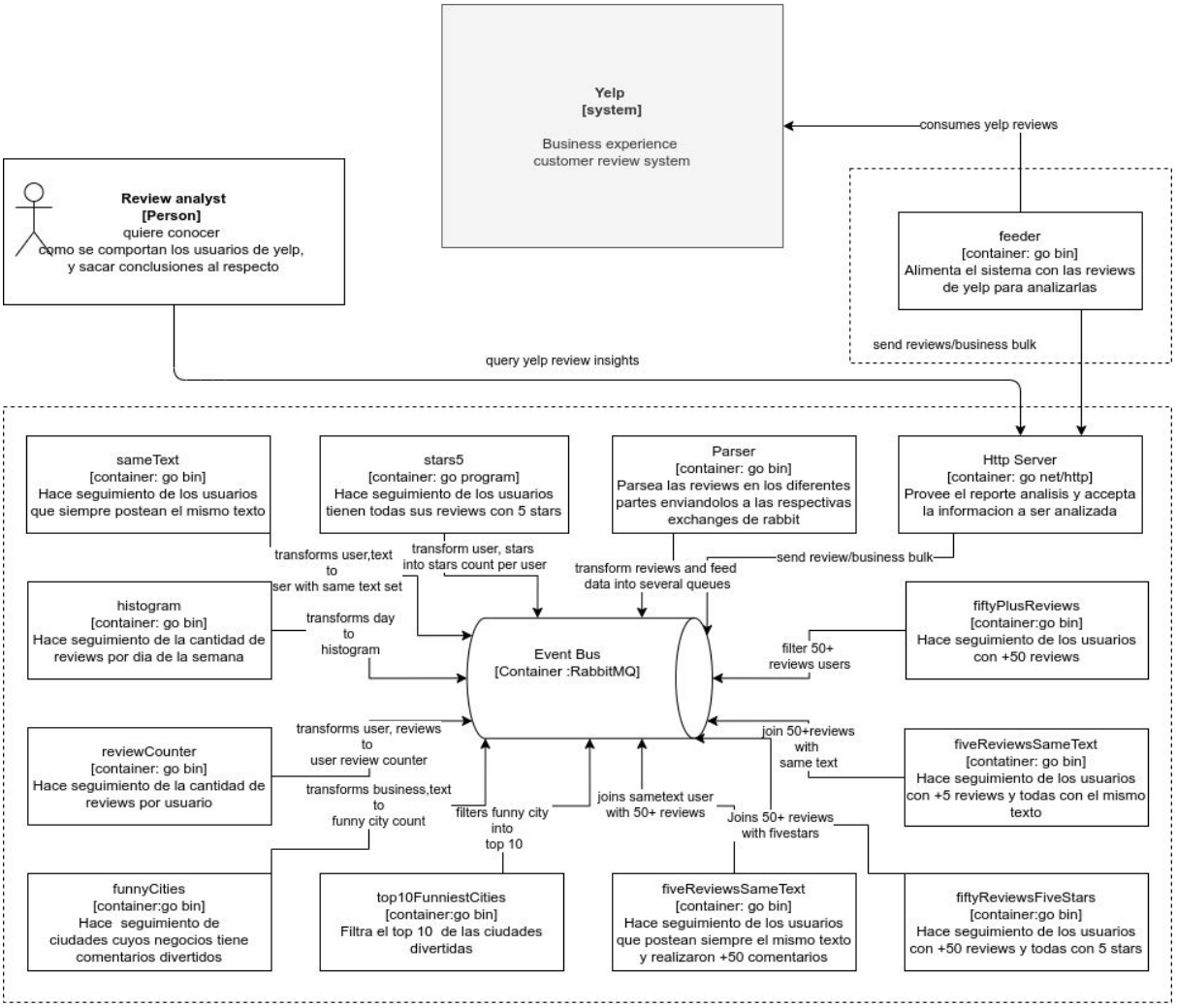
System Context Diagram for yelp reviews analysis system

La idea del sistema es que una persona interesada en sacar conclusiones sobre los usuarios de yelp consulte el sistema.
Idealmente el sistema de alimentaria en vivo con reviews y businesses de yelp.

El procesamiento de las reviews sigue el siguiente DAG.



Containers



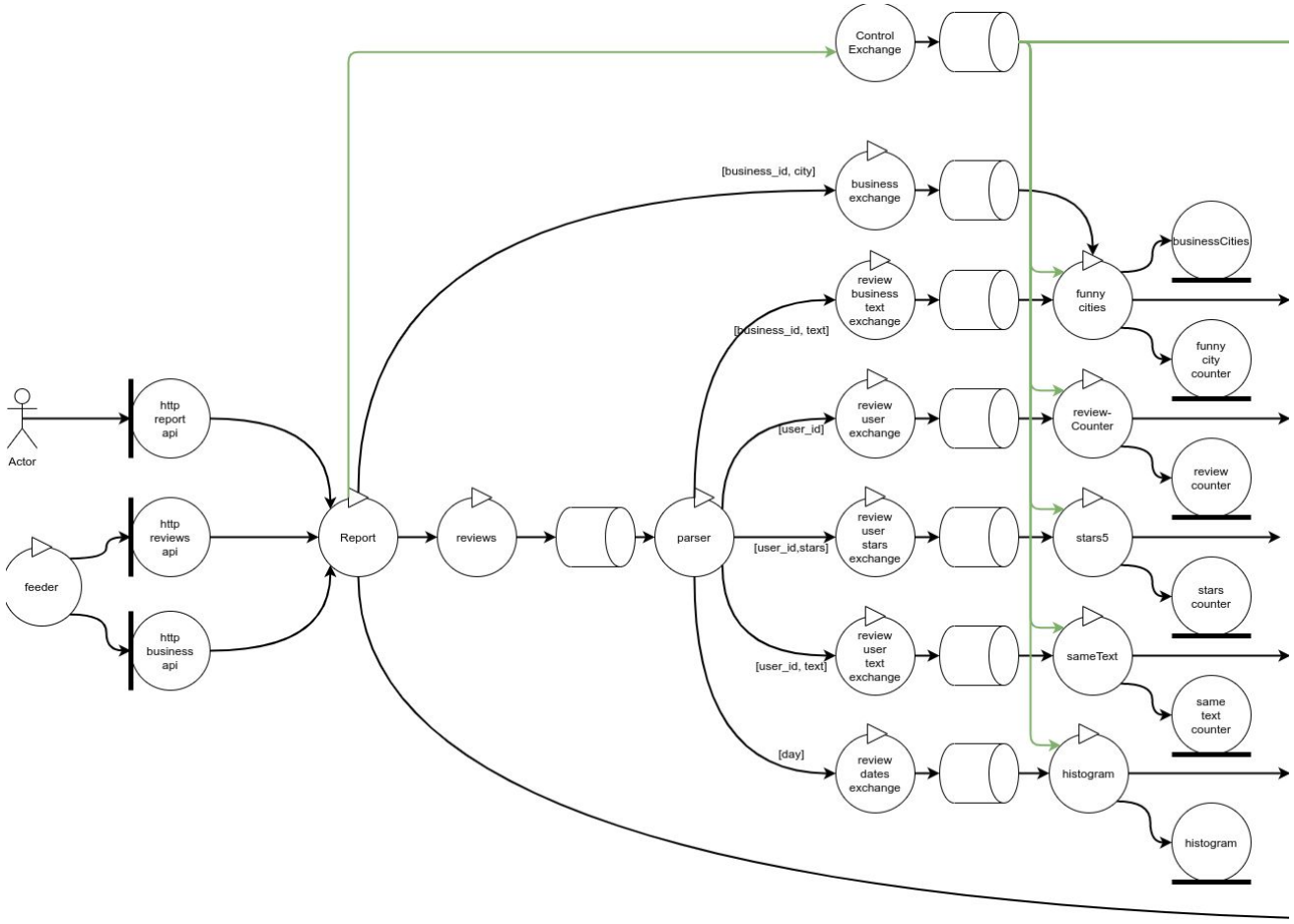
Container Diagram for yelp reviews analysis system

El sistema está compuesto de varios contenedores que realizan alguno de los pasos del DAG anterior.
Todo contendores se comunican a través de colas de rabbit.
Y el contendor httpServer/report es quien recibe los request para processar reviews, business y también para consultar el reporte.

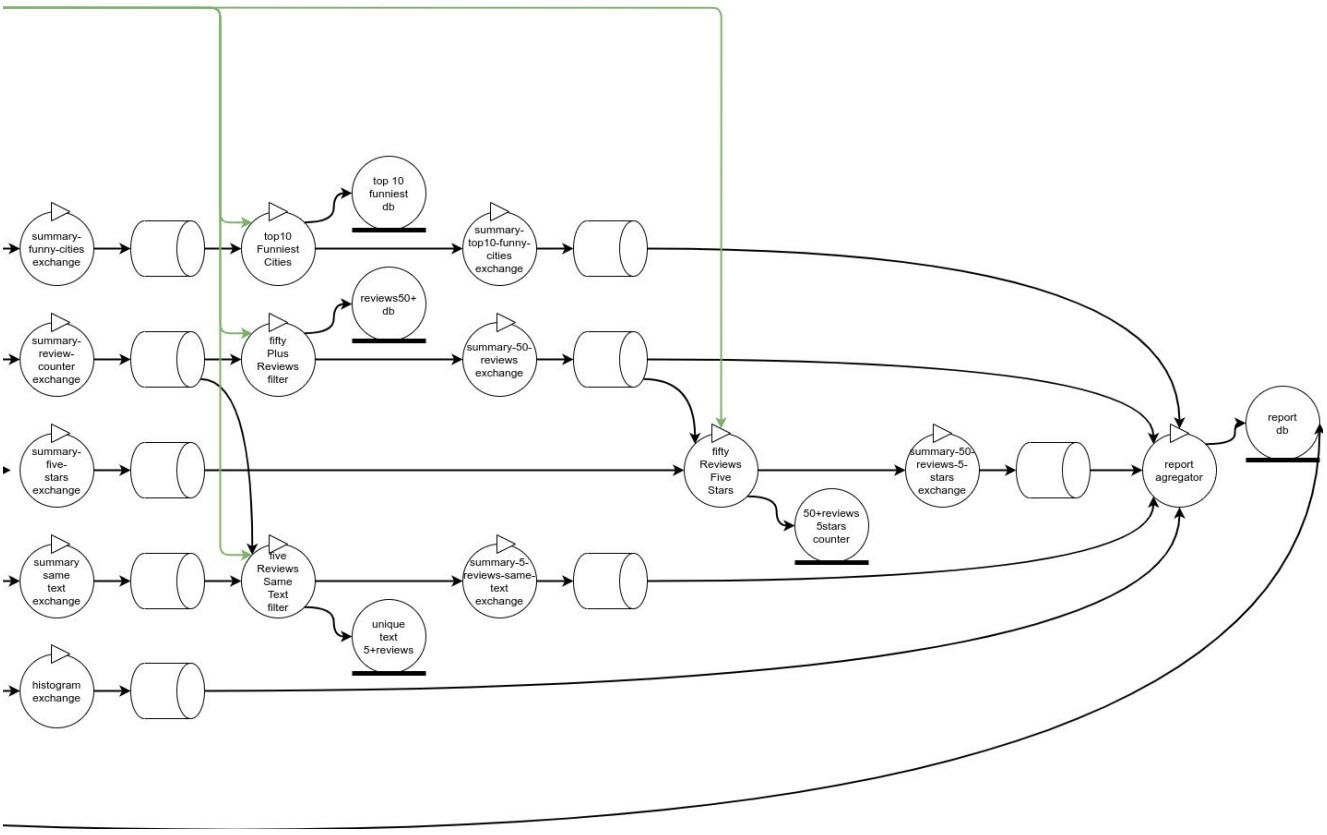
En diagrama de robustez vemos que hay un exchange de rabbit dedicado a control de los contenedores. Todo los contenedores que se comuniquen con este exchange crean una cola anónima para implementar pub/sub sobre este exchange.

En particular el contenedor de reportes verifica mediante este exchange que todos los contenedores estén activos antes de devolver un reporte.

En el diagram de robustez vemos que cada exchange(boundary) tiene su propia cola.
En particular todos los exchange de datos(reviews/business) que tiene un único consumidor, tiene una queue nombrada.
Los exchange que contienen datos que van a ser consumidos por más de un contenedor, declaran queues anónimas para poder implementar pub/sub en estos casos.

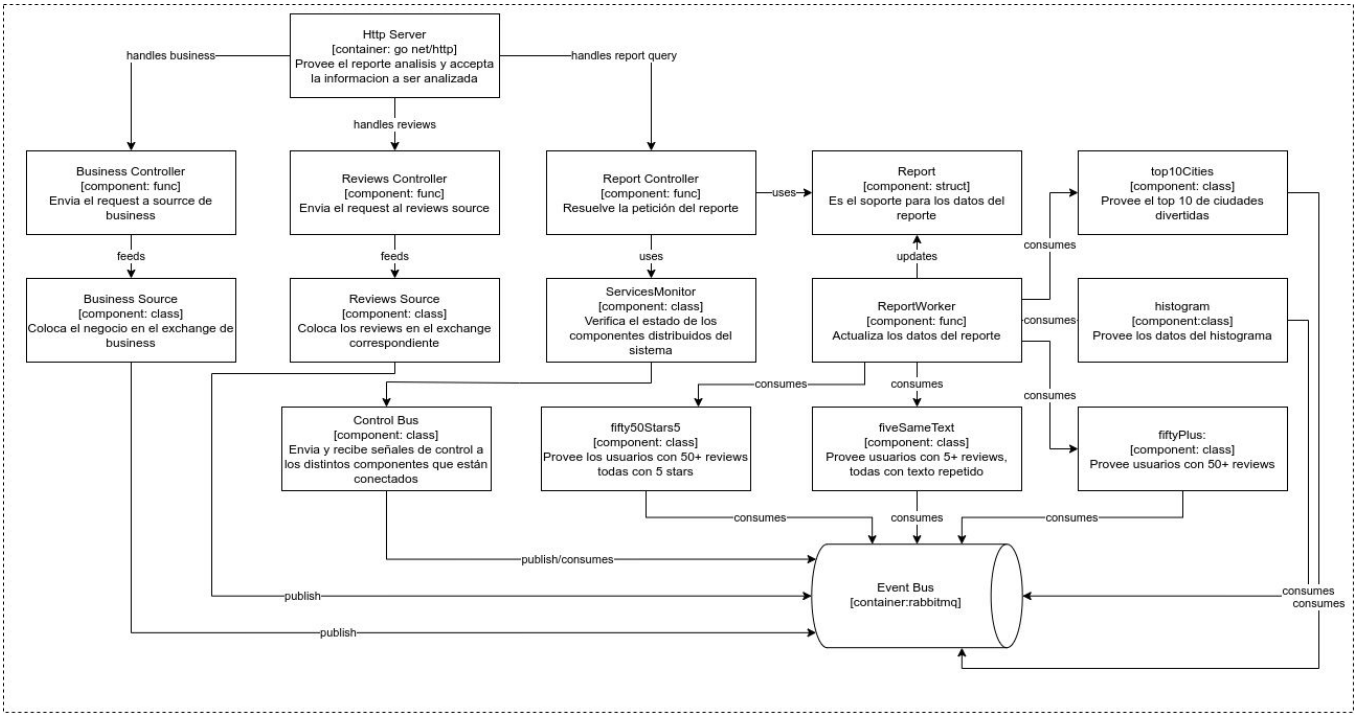


....


summary-funny-cities exchange
top10 Funniest Cities
top 10 funniest db
summary-top10-funny-cities exchange
summary-review-counter exchange
fifty Plus Reviews filter
reviews50+ db
summary-50-reviews exchange
summary-five-stars exchange
fifty Reviews Five Stars
50+reviews 5stars counter
summary-50-reviews-5-stars exchange
report agregator
report db
summary same text exchange
five Reviews Same Text filter
unique text 5+reviews
summary-5-reviews-same-text exchange
histogram exchange

Components

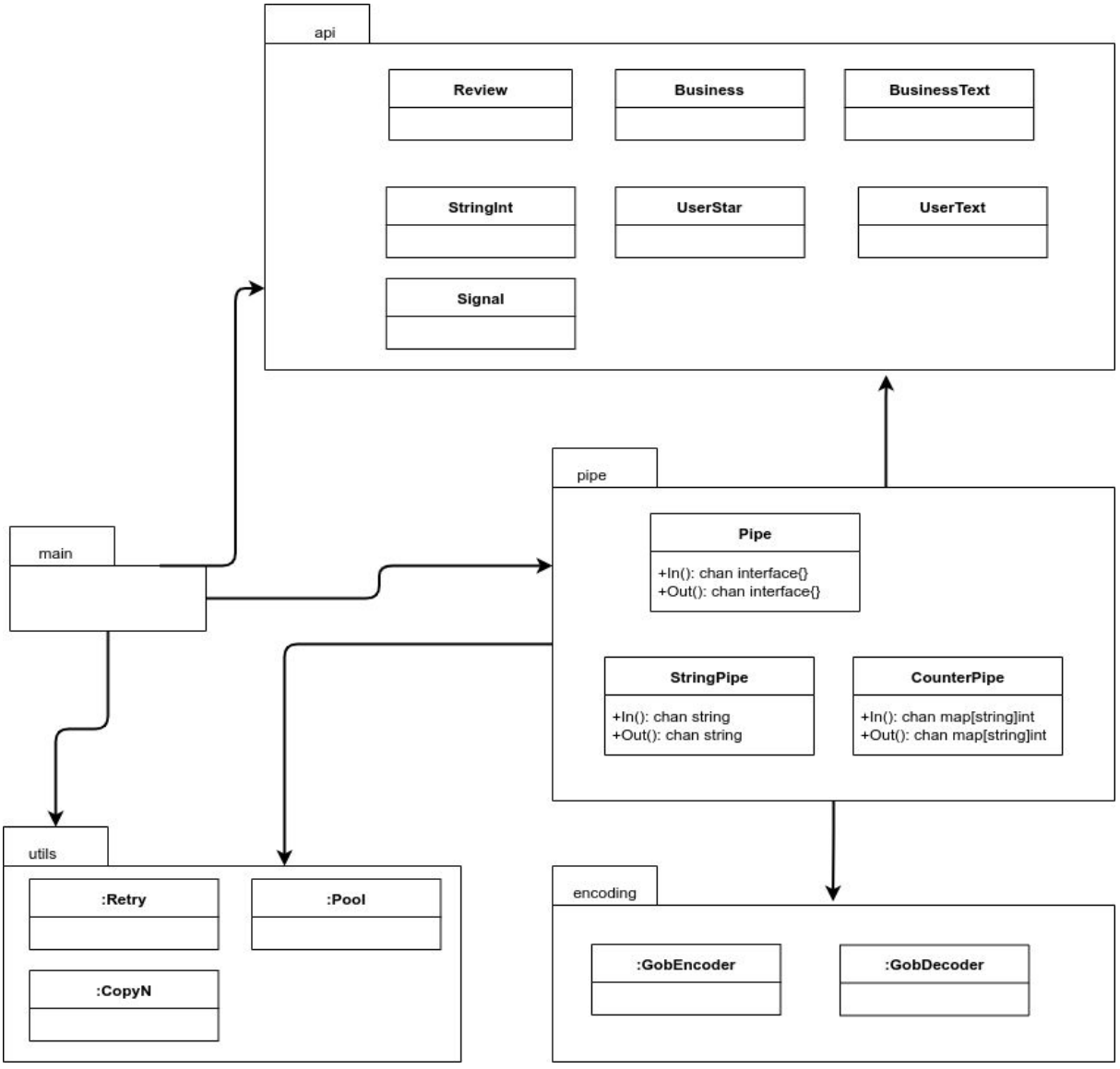
Cada contenedores es se conecta a las correspondientes queues de rabbit y consume/produce según corresponda.
En particular el httpServer/report es el contenedor más complejo y además de consumir las queues correspondientes para armar el reporte final, atiende request http y verifica utilizando el exchange control que el resto de los contenedores están habilitados para procesar datos.



Component Diagram Report Server for yelp reviews analysis system

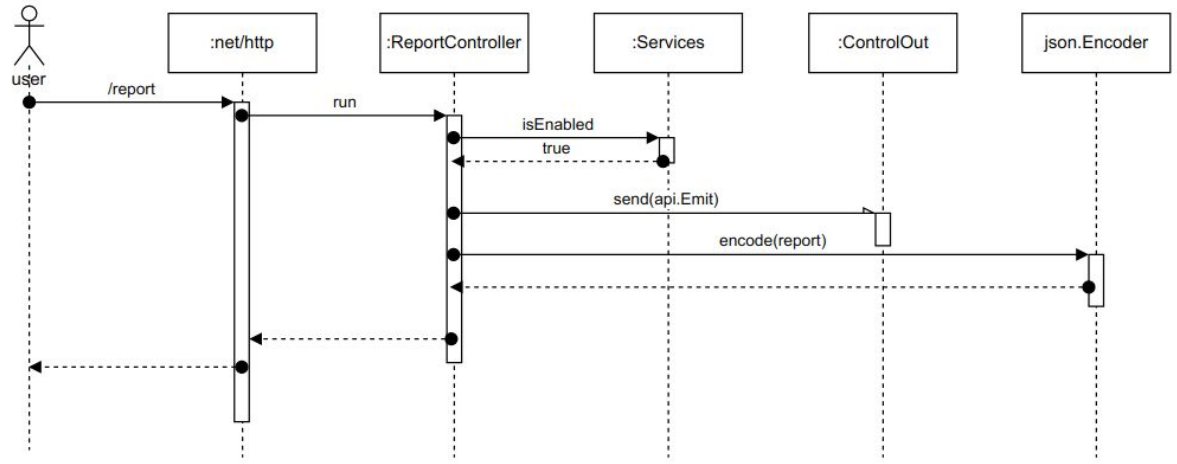
Code

Las interacciones con rabbit están modeladas como un “pipe”, que se puede ver en package pipe.
Todo los pipes que utiliza el sistema están declarados como estáticos.
Los pipe tienen una punta de entrada y una de salida.
Cuando un componente quiere conectarse a una pipe en particular, invoca el método In,Out (que encapsula la instanciación de conecciones y channels al rabbit que correspondan) y obtiene un channel de go para enviar mensajes a través de ese pipe.

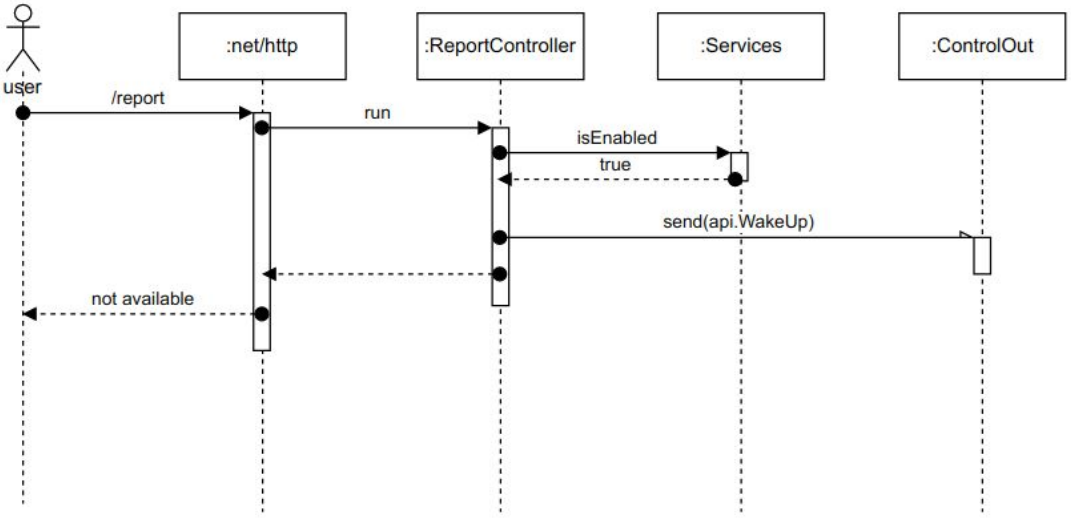


Los mensajes que se envian a rabbit se wrappean en Message y se serializan utilizando gob, esto esta encapsulado por el package encoding.

Los contenedores en general van a obtener los pipes que necesiten y se van a suscribir al pipe de control. Realizando un select sobre el pipe de control y pipe de datos que corresponda.
En caso de recibir datos, actualizan el estado que mantiene ese contenedor.
En caso de recibir una señal por el pipe de control, va a reaccionar acorde a la señal.
Si es wakeup van a responder con un join al pipe de control.
Si es emit, van a emitir la información de estado que mantiene.
A continuación vemos el diagrama de secuencia del httpserver/report.
El mismo utiliza una clases Services que se encarga de revisar que todos los contenedores estén habilitados.



Sequence Diagram For Services Ready



Sequence Diagram For Services Not Ready

Comentarios

TODOs:

- Graceful Shutdown:
    - Si bien está la idea a medio implementar de orquesta un shutdown de todo el sistema, hay algunos contenedores que utilizan un pool de go rutinas y sólo uno obtiene la señal de quit.
    - No se está manejando la señal de kill en los contendores, por lo que en caso de abortar, no se hace apropiadamente.