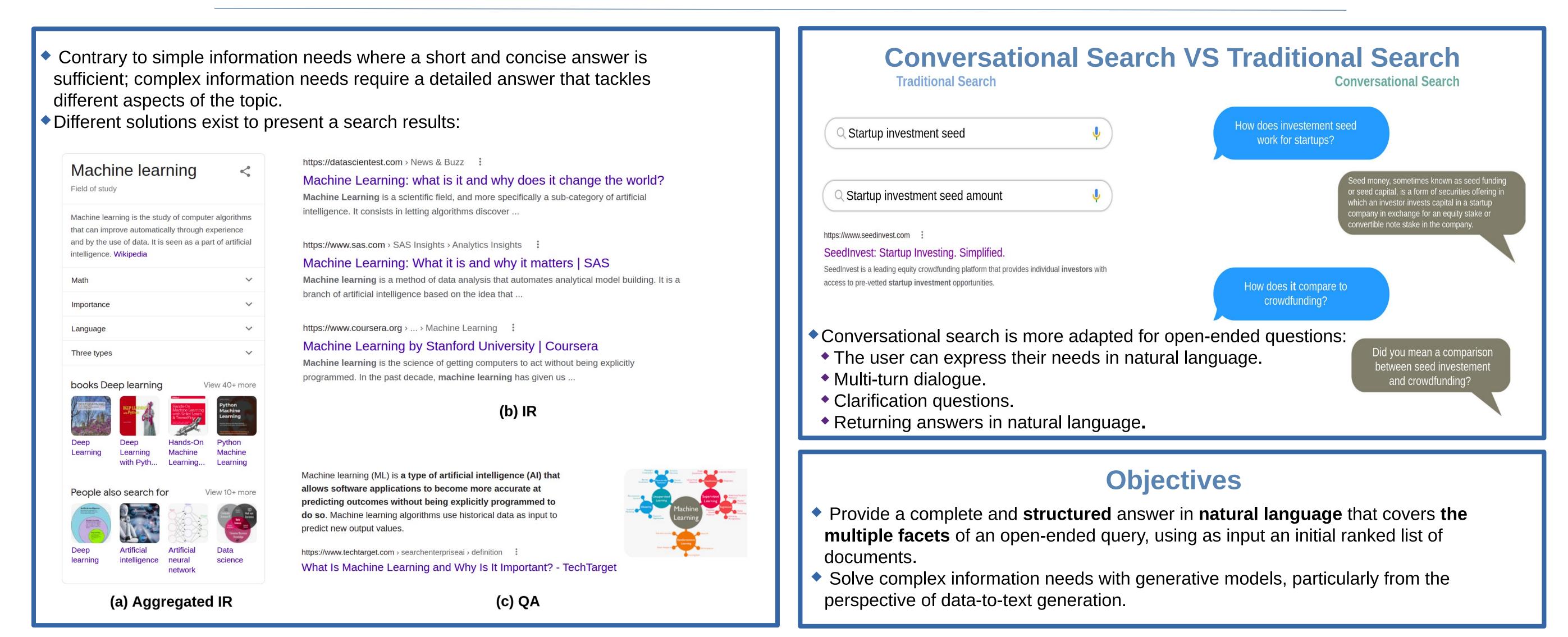




# **Does Structure Matter? Leveraging Data-to-Text Generation for Answering Complex Information Needs**

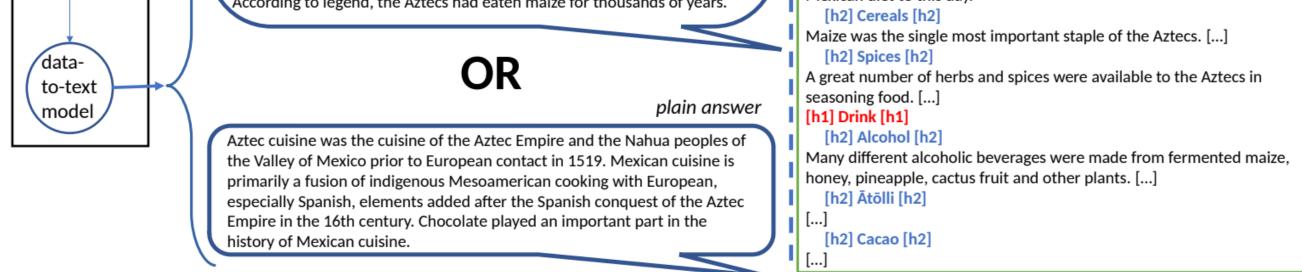
Hanane Djeddal, Thomas Gerald, Laure Soulier. Karen Pinel-Sauvagnat, Lynda Tamine Sorbonne Université, CNRS, Institut des Systèmes Intelligents et de Robotique (ISIR), in collaboration with Institut de Recherche en Informatique de Toulouse (IRIT)

## **Context, Motivation and Objectives**

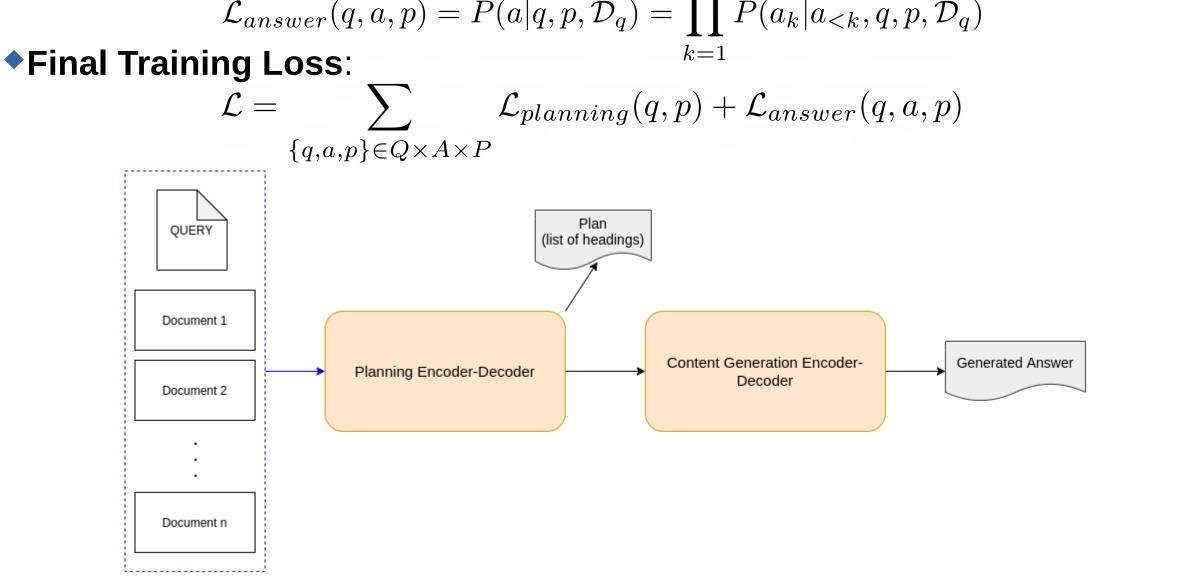


## **A Data-to-Text Approach for Answer Generation**

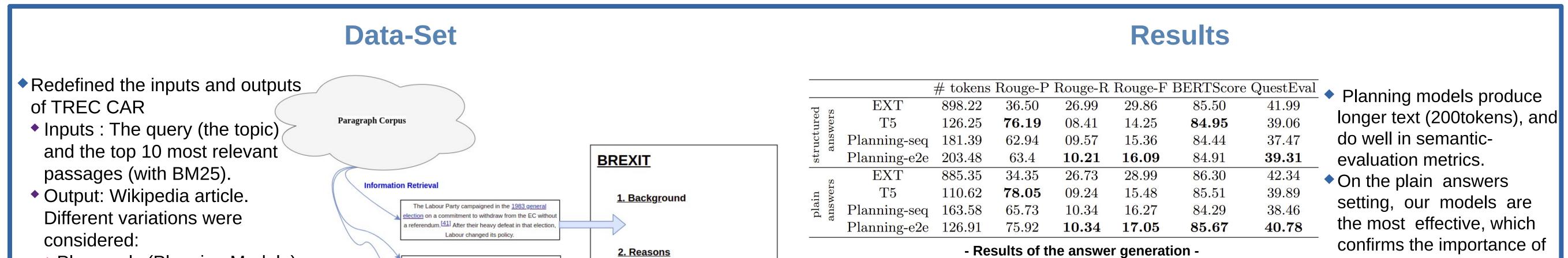
| query         |  |   | <b>Content Selection and Planning Pipeline</b>   |
|---------------|--|---|--|
| aztec cuisine | <pre>structured answer aztec cuisine was the cuisine of the Aztec Empire and the Nahua peoples of the Valley of Mexico prior to European contact in 1519. [h1] Etymology [h1] The word xocolatl is derived from the Nahuatl word xocolatl. [h1] Aztec cuisine [h1] [h2] Mexican food [h2] Mexican cuisine is primarily a fusion of indigenous Mesoamerican cooking with European, especially Spanish, elements added after the Spanish conquest of the Aztec Empire in the 16th century. [h2] Chocolate [h2] Maize many other types of maize that were introduced by the Aztecs. [h2] Other foods [h2] There are many other types of maize that were introduced by the Aztecs. [h1] History [h1] According to legend, the Aztecs had eaten maize for thousands of years.</pre> | Aztec cuisine was the cuisine of the Aztec Empire and the Nahua peoples of the Valley of Mexico prior to European contact in 1519. []         [h1] Meals [h1]         Most sources describe two meals per day, though there is an account of laborers getting three meals, one at dawn, another one at around 9 []         [h2] Feasts [h2]         Many accounts exist of Aztec feasts and banquets and the ceremony that surrounded them. []         [h1] Food preparation [h1]         The main method of preparation was boiling or steaming in two-handled clay pots or jars called xoctli in Nahuatl and translated into Spanish as olla ("pot"). []         [h1] Foods [h1]         The Aztec staple foods included maize, beans and squash to which were often added chilis, nopales and tomatoes, all prominent parts of the Mexican diet to this day.         [h2] Cereals [h2] | <ul> <li>◆Planning encoder-decoder: Encodes each document d<sub>q</sub> ∈ D<sub>q</sub> concatenated with the query q and decodes a plan p.</li> <li><i>L</i><sub>planning</sub>(q, p) = P(p q, D<sub>q</sub>) = ∏<sub>j=1</sub><sup> p   h<sub>j</sub> </sup> P(h<sub>jk</sub> h<sub>j,<k< sub="">, q, D<sub>q</sub>)</k<></sub></li> <li>◆Content generation encoder-decoder: Encodes each heading h<sub>p</sub> in the plan p, concatenated with the embedding of the document list D<sub>q</sub> and decodes an answer a.</li> <li><i>L</i><sub>answer</sub>(q, a, p) = P(a q, p, D<sub>q</sub>) = ∏ P(a<sub>k</sub> a<sub><k< sub="">, q, p, D<sub>q</sub>)</k<></sub></li> </ul> |



- Data-to-Text Generation introduces the notion of structure at two levels:
- At input level (encoding): exploiting the structure of the inputs (cells, rows, etc)
- At output level (decoding): structuring the output text by means of content selection and planning.



## **Evaluation Setup & Results**



### Plans only (Planning Module) Structured or plain answer (Content Generation Module)

#### February 2019, the parliamentary Digital, Culture, Media and Sport Committee called for an inquiry into "foreign fluence, disinformation, funding, voter manipulation, and the sharing of data" in the Brexit vote 3. Impact Many effects of Brexit depended on whether the UK left vith a withdrawal agreement, or before an agreement was ratified ("no-deal" Brexit)

- TREC CAR DATA-SET

[h2] Cocoa pod [h2] A cocoa pod (fruit) has a rough, leathery rind about thick (this varies with the origin and variety of pod)filled with sweet, mucilaginous pulp (called baba de cacao in South America) with a lemonade-like taste enclosing 30 to 50 large seeds that are fairly soft and a pale lavender to dark brownish purple color.

#### [h2] Varieties [h2]

[h1] Production [h1]

The three main varieties of cocoa plant are Forastero, Criollo, and Trinitario. [h2] Harvesting [h2]

Cocoa trees grow in hot, rainy tropical areas within 20° of latitude from the Equator. Immature cocoapods have a variety of colours, but most often are green, red, or purple, and as they mature, their colourtends towards yellow or orange, particularly in the creases. [h2] Harvest processing [h2]

The harvested pods are opened, typically with a machete, to expose the beans. A typical pod contains 20 to50 beans and about 400 dried beans are required to make one pound - or 880 per kilogram - of chocolate. The wet beans are then transported to a facility so they can be fermented and dried. The beans should bedry for shipment (usually by sea). Throughout Mesoamerica where they are native, cocoa beans are used for avariety of foods.

#### [h2] World production [h2]

Nearly of cocoa are produced annually. The historical global production was : 1974 : 1,556,484 tons, 1984 :1,810,611 tons, 1994 : 2,672,173 tons, 2004 : 3,607,052 tons, 2013 : 4,585,552 tons. The production increased33 by 194.6% in 39 years, representing a compound annual growth rate of 2.81%. About 4.585 million

tonnes of cocoa beans were produced in the 2013â2014 growing year, which runs from October to

Funded by:

#### [h2] Child slavery [h2]

The first allegations that child slavery is used in cocoa production appeared in 1998

#### - STRUCTURED ANSWER -

A cocoa pod (fruit) has a rough, leathery rind about thick (this varies with the origin and variety of

pod)filled with sweet, mucilaginous pulp (called baba de cacao in South America) with a lemonade like taste enclosing 30 to 50 large seeds that are fairly soft and a pale lavender to dark brownish purple color.

The three main varieties of cocoa plant are Forastero, Criollo, and Trinitario. Cocoa trees grow in hot, rainy tropical areas within 20° of latitude from the Equator. Immature cocoapods have a variety of colours, but most often are green, red, or purple, and as the mature, their colourtends towards yellow or orange, particularly in the creases.

The harvested pods are opened, typically with a machete, to expose the beans. A typical p contains 20 to 50 beans and about 400 dried beans are required to make one pound - or 880 per kilogram - of chocolate. The wet beans are then transported to a facility so they can be fermented and dried. The beans should bedry for shipment (usually by sea). Throughout Mesoamerica wher they are native, cocoa beans are used for avariety of foods

Nearly of cocoa are produced annually. The historical global production was : 1974 : 1,556,484 tons, 1984 :1,810,611 tons, 1994 : 2,672,173 tons, 2004 : 3,607,052 tons, 2013 : 4,585,552 tons. The production increased33 by 194.6% in 39 years, representing a compound annual growth rate of 2.81%. About 4.585 milli tonnes of cocoa beans were produced in the 2013â2014 growing year, which runs from October t Septembe

The first allegations that child slavery is used in cocoa production appeared in 1998

#### - PLAIN ANSWER

[h1] Production [h1] [h2] Harvest processing [h2] [h2] World production [h2]

#tokens #heading depth Rouge-P Rouge-R Rouge-F BERTScore Meteor  $\overline{\mathrm{T5}}$ 1.14 **39.89** 77.402.2407.693.24FP 1.41 04.695.97ĪĒ 1.45 31.20 8.29 11.51 $\overline{81.25}$ 1.834.42 $\frac{1}{\text{Planning-seq}} \stackrel{--}{\text{FP}}$ 1.88**1.45** 31.31 7.9311.034.1180.495.555.51 $3.\overline{3}7$  $\overline{8}1.\overline{2}1$  $\overline{1.15}$   $\overline{35.15}$ 07.3411.121.57IP Planning-e2e 3.271.1634.7906.3809.7880.704.71

### - Analysis of the intermediate and final plans -

explicitly present in the final output. Our plans cover more facets, in correct order with a better relevant semantics.

structure prior even if it's not

## Conclusion

Generating a complex answer in a single turn to open-ended queries proves to be challenging because of the length of both inputs and outputs. Modeling a structure prior is beneficial to guiding the final output generation.

Data-to-Text generation approaches provide a promising framework that can be applied to this task.

### <u>{hanane.djeddal,thomas.gerald,laure.soulier}@isir.upmc.fr</u> <u>{sauvagnat,Lynda.Lechani}@irit.fr</u>



Under joint supervision



[h2] Cocoa pod [h2] [h2] Varieties [h2] [h2] Harvesting [h2] [h2] Child slavery [h2]

- PLAN ONLY -