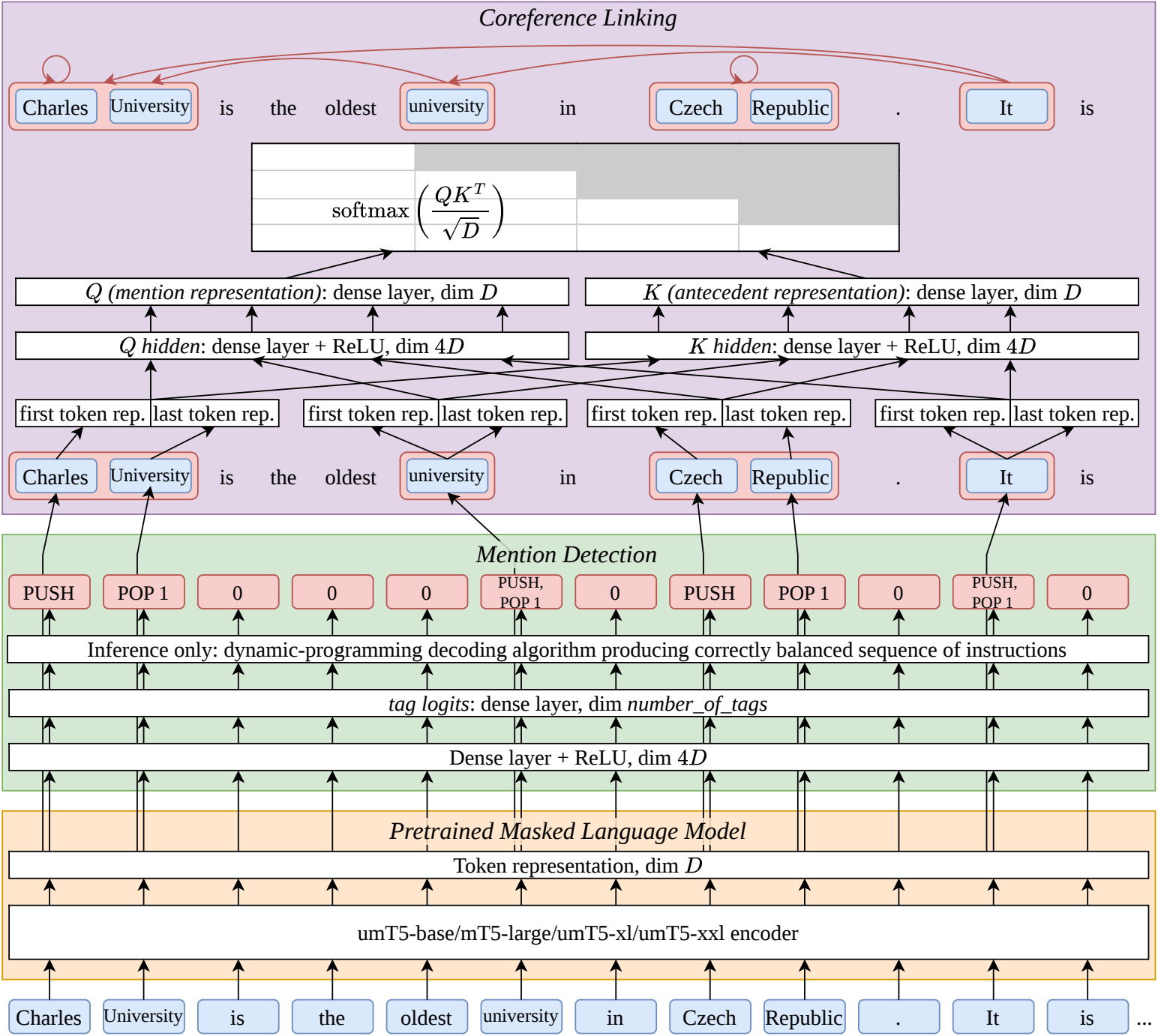
Coreference Linking
Charles
University
is
the
oldest
university
in
Czech
Republic
.
It
is
softmax $\left(\frac{QK^T}{\sqrt{D}}\right)$
Q (mention representation): dense layer, dim D
K (antecedent representation): dense layer, dim D
Q hidden: dense layer + ReLU, dim $4D$
K hidden: dense layer + ReLU, dim $4D$
first token rep.
last token rep.
Mention Detection
PUSH
POP 1
0
PUSH, POP 1
Inference only: dynamic-programming decoding algorithm producing correctly balanced sequence of instructions
tag logits: dense layer, dim *number_of_tags*
Dense layer + ReLU, dim $4D$
Pretrained Masked Language Model
Token representation, dim D
umT5-base/mT5-large/umT5-xl/umT5-xxl encoder
...