这篇文献的名字叫做Hierarchical Transformers for KGE。在这篇论文里他提出了一个在两个不同层面上做Transformer的KGE方法。首先他总结了一下过去的KGE方法，传统的KGE方法是在向量空间里设计一些几何规则实现实体和关系之间的匹配，如TransE，DisMult等，它们在概念上容易被理解，但是没有明显的使用到结点周围的环境信息。为了考虑环境信息，随后出现了使用GNN或者注意力机制的方法，但是使用GNN类方法的话模型效果会受到网络层数的限制，所以在这里他提出了使用Transformer做KGE，并且是多个层面的Transformer。

首先，它第一个层面Transformer是实体层面的Transformer，输入的是一个单独的实体和它的关系，这个实体可以是头实体或者尾实体，以及一个特殊的tokenCLS，经过Transformer得到这个特殊Token的输出，简单的Transformer方法的话可能会直接把这个特殊token的表示拿到下游任务里做一个链接预测。但是作者认为这样一个简单的Transformer只考虑了单个实体和它对应的关系之间的信息，而没有考虑周围的环境信息，因此在得到每个entity和relation pair的embedding之后，它将某个实体和它周围邻居的pair embedding输入到第二个层面的Transformer，它称之为环境层面的Transformer。然后得到环境层面的transformer的输出作为最终的输出，然后使用这个输出进行连接预测，具体来说就是去计算这个embed ing和训练集中的target embedding的分数，以及其他实体embedding和target embedding之间的分数，对他们做一个sofemax之后再做一个交叉熵损失。

接着他还对输入的一部分实体做了mask操作，然后对希望transformer能够对这个掩码进行还原，并把这部分的损失也加到了损失函数里。

在实验部分的话，它使用了FB15K-237和WN18RR两个数据集，效果还是挺好的。

然后在最后面他还做了一个QA的任务，因为他认为Transformer结构之间是具有可迁移性的，所以使用Bert做QA的话，可以把Transformer里的知识也逐层的加进去，不过对于QA的任务我目前不是很熟悉。就是说，它的查询来自于原始的BERT，键值来自于它训练出的Transformer