7-6. 了解目前常见的几种汉语词性标注集，比较它们的差异，并阐述你个人的观点。

这里是一些常见的词性标注集：

1. **北大计算所词性标注集（PKU POS Tagging Set）：**该词性标注集广泛用于中文自然语言处理任务，包括名词、动词、形容词、副词、数词、量词、介词等多种词性。
2. **863词性标注集：**这是中国“863计划”自然语言处理项目中采用的词性标注体系，包括基本词性和细分词性。基本词性包括名词、动词、形容词、副词等，细分词性包括地名、人名、机构名等。
3. **中科院计算所（ICT）词性标注集：**这是中科院计算技术研究所发布的一种词性标注体系，包括名词、动词、形容词、副词、数词、量词、介词等多种词性，以及一些细分词性。
4. **国标词性标注集（GB/T 20542-2006）：**这是中国国家标准化管理委员会发布的一种词性标注体系，该体系中包括名词、动词、形容词、副词、数词、量词、介词等多种词性，以及一些细分词性。

PKU这个词性标注集我看到很多项目都在用，比较经典，包含36个词性类别。很多开源工具比如jieba参照的都是这个词性标注集。

863的标注集词性类别比较少，只有25个，可能比较适合精简一点的情况。我看到有HanLP之类的平台在用它。

ICT的词性标注集我看到有不少开源项目在用它。这也是ICTCLAS使用的词性标注集，说是比较适合工程应用，有42个词性类别。

GB/T 20542-2006是国标词性标注集，类别很多，有56个。比较权威和通用。不过这份标准有些老了，相比于很多大学推出的词性标注集能够不断迭代，这份可能不太适合网络用语的标注。我好像没看到太多的开源项目在用这个，可能是因为词性太多了吧。

然后除了这几个比较传统的，各大国内公司使用的都是自己根据已有标注集调整过的词性标注集，我看到百度、阿里、boson等标注集都不是完全遵照常见标准的，应该都是自己调整过的。

感觉似乎在中文NLP领域词性标注集的标准还是挺混乱的，各家都有自己的标准，要训练中文NLP任务如果用了多个基于不同词性标注集的语料库，还需要手动将它们统一一下才好处理。