AITest Doc

Aim：通过采集开发人员和测试人员在开发、测试的初步阶段，对网页点击验证（实际是对具体控件），收集一定规模的数据量，作为训练集，训练一定的算法（聚类、复杂网络重要节点分析），寻找到网页空间的关键节点和优化路径，以此为依据，将训练得到的数据，交给自动化测试工具，以便更加快捷高效地完成网页的自动化测试。

Steps：

1. 收集数据：利用网页点击插件（eg.UI Recorder ，Sideex）在点击页面时，插件收集数据，形成训练集，数据的内容格式以json串的形式，在点击完成后，存在server的MangoDB1里（HanRui）
2. 清洗数据：从MangoDB1里取到数据，根据算法需求，清洗数据，以key-value的数据格式，其中在算法过程中需要标记唯一的标识控件的信息ID,初步ID=target+pages（待确认），以此ID来唯一回溯到控件信息（WangShuai&Shuanbao）
3. 建模分析：
4. 利用聚类算法DBSCAN和复杂网络重要节点排序算法，寻找Hub节点及测试路径，返回结果需要携带控件ID信息，存入新的Database.MangoDB2里。
5. 自动化测试（XiaoMin），从MangoDB2里取到数据，以ID信息的排序，依次进行测试
6. 结果分析：对于算法输出的携带ID信息的路径，使用自动化测试能否顺利通过测试，作为一个重要指标。

另外如果案例都可以通过，那么我们就需要寻找到如何评价算法的好与坏的指标。（进一步的目标）

1. 验证结果：对于每一个测试路径，需要作出测试通过标记，测试人员通过分析测试路径标记，确保所有测试案例都已经通过测试。最终交付测试完成的产品（网页）

算法的issue: (HanRui 、Wangshuai & Shuanbao)

a. 逻辑性要理清楚，比如我们拿到训练集的数据，进行运算后，再用测试集来验证，每一次运算是使用单个页面的点击量还是所有页面的点击量，进行排序，区别在哪里呢？

b. 页面和页面之间如有耦合关系，如何解耦？

c. 对于我们而言，点击量就是意味着点击网页背后控件的重要程度，那么对于小数据样本，噪声如何消除？如果加权，如何赋权重？

Tools：

1. 图形化编程 Sikuli-X http://www.sikuli.org/
2. 网页采集数据控件Sideex <http://www.sideex.org/>
3. UI Recorder https://github.com/alibaba/uirecorder
4. Selenium <http://docs.seleniumhq.org/>
5. MangoDB