Relevance Project Framework

Process Guidelines for Conducting a Predictive Coding Project with the Equivio Relevance Application
This document offers procedural guidelines for the conduct of a predictive coding project with the Equivio Relevance application. The document describes the various stages of the process flow. The document is intended as input for the use of consultants who are responsible for designing predictive coding project protocols, as well as for Equivio internal and client project leaders responsible for managing such projects.

The document does not detail roles and responsibilities for the different functionaries that form the project team, such as IT personnel, litigation support specialists and attorneys. Similarly, the scope of the document does not extend to IT functions such as back-up and database administration.

While the document does provide detailed guidelines for managing and standardizing the predictive coding process flow, new users are encouraged to consult with Equivio, especially at critical decision points, such as the completion of training and QA procedures.

Following preliminary preparation activities, the process breaks down into the following phases: assessment, training, catch-up (special case of training), decision and verification.

**PREPARATION**

1. The first step is to prepare the data. This includes data extract and building of a search index.
2. Metadata culling should be performed prior to loading the data into the predictive coding application (Relevance). This stage includes de-duping, de-NISTing, and culling by date range, file types and custodian. In rare cases, basic keyword filters will also be applied. It is recommended that application of keyword filters be minimized as keyword filters are liable to inadvertently cull significant sections of relevant data.
3. Prior to starting the process, it is recommended to engage with opposing counsel, communicating intention to use predictive coding, and verifying the definition of relevance to be applied.
4. In this protocol, the term "expert" is used to refer to the person training the Relevance system. It is important to select the expert with care. Relevance is a "garbage-in garbage-out" application. In other words, the quality of the software’s output is highly dependent on the quality of the human’s input. While the system does offer internal checks to verify the consistency of the expert’s input tags, it is not able to detect consistent systematic errors. The expert should be selected with this in mind. As such, the expert should be an experienced and knowledgeable attorney, whose judgment can be relied upon, and who is very familiar with the case. To the extent that the expert is new to the case, it is recommended that the team engage in discussions to hammer out the definition
of relevance, as well as *ad hoc* searches and review of the data to familiarize herself with the data.

NOTE: If there are multiple issues, there may be more than one expert, with each expert training a number of issues. However, for any given issue, the system should be trained by only one expert.

**ASSESSMENT**

1. In order to ensure statistical validity of the results, it is critical to separate the "control" documents from the "training" documents. For any matter, the classifier, which calculates the relevance score for documents in this particular case, is created by the training documents. The performance of the classifier is measured against the control documents.

2. The control set is created in the "assessment" phase. Assessment is the first phase of the Relevance process. These control documents represent the "gold standard" against which the progress of training is monitored, the performance of the classifier is checked, and results are quantified.

3. In terms of statistics generated, the system uses the control documents to estimate the richness of the collection, and the recall and precision achieved by the classifier.

4. The Assessment phase is a two-stage process:
   a. The first set of control documents comprises 500 documents, selected randomly by the Relevance system. The "expert" tags these sample documents as Relevant or Not Relevant.
   b. The system uses these first 500 documents to generate an initial estimate of richness. In addition, the system generates an error margin for the estimate of recall. The error margin calculation for recall depends on three main factors:
      - Number of relevant documents in the control set: For example, 62 relevant documents yield an error margin of +/-10%, while 246 relevant documents yield an error margin of +/-5%.
      - Confidence level: 95%. This is the standard confidence level used in most statistical applications.
      - Target recall level: 80%. This parameter refers to the expected minimal recall level that will be achieved by the system, after training. Note that the worst-case error margin (that is, the largest error margin) occurs given a target recall level of 50%.

---

1 The calculations also assume a very large population (over 2,000,000 documents). This is a worst-case assumption, because for smaller populations, slightly smaller samples are required. Beyond the assumed size, the impact of population size on sample requirements is negligible.

2 For example, given a control set that contains 62 relevant documents, the error margin for recall of 80% is +/-10%, while for recall of 50% the error margin is over +/-12.5%. The corollary also holds: assuming recall
historical use of the system shows that in virtually all cases, users will select the document review set to yield at least 80% recall.

c. Following tagging of the first 500 control documents, the system presents the current status, and allows the user to adjust these parameters. For example, the first 500 documents may have yielded a recall error margin of +/-13.25% (this is the error margin produced by 35 relevant documents). Retaining the default assumptions for confidence level and target recall, the user might decide to set a target error margin of 10%. This error margin will require additional 385 control documents.

5. It is recommended that the Assessment phase, and at least the first 500 documents, be conducted by a group of two or three attorneys. Under this approach, the group is required to reach consensus on the tag for each control document. This collaborative approach helps ensure that the notion and scope of relevance is clearly bounded and well defined.

6. If and where required, the relevant control documents from the control set will be disclosed to the requesting party. Use the system’s near-duplicate groupings and email threads to verify the consistency of the Assessment tags. Be sure not to disclose non-relevant or privileged data.

TRAINING

Once the Assessment phase is complete, with the required validation level, begin the training process. Training is an iterative process, as follows:

1. The system selects a sample of 40 documents. The first sample is a random sample. Subsequent samples are selected using an Active Learning approach. Under Active Learning, each training sample is selected based on what has been learnt from previous samples. In selecting sample documents, the system’s objective is to maximize the sample’s contribution to the training process; in other words, to choose a sample that will teach the system as much as possible about the population of documents. Based on this criterion, the system chooses samples that provide comprehensive coverage of the population (reducing under-inclusiveness), while also fine-tuning and nuancing the concept of relevance that the classifier has developed to date (reducing over-inclusiveness). This multi-layered approach ensures that the classifier minimizes its exposure to a cross-section of relevance documents that is as broad as possible, and in so doing,
broaden its concept of relevance to capture more of the relevant documents, while, on the other hand, refining the classifier to eliminate false positives.

NOTE: In most cases, seeding is not used and is not required. Seeding refers to a situation where the user has a set of documents which he knows to be relevant, and which can be fed into Relevance to train the system. Unaided seeding is liable to bias the results; the system will find only documents which are similar to the seed documents, but will not capture other types of relevant documents which may be present, unbeknown to the application, in the population. Relevance supports the input of seed documents, and is built to counterbalance the self-fulfilling prophecy trap caused by the seed bias. However, it is recommended to limit the use of seeding to situations where the richness of the population is very low. In such cases, seeding is used to accelerate the initial training process. Indeed, in order to obtain full benefit from the seed documents, it is recommended to perform seeding at the outset of the training process, immediately following assessment. Following the seed input, it is mandatory to conduct the full training cycle, with the standard 25-40 rounds. This is critical in order to re-dress the potential skew effect of the seed documents.

2. The expert tags the sample documents as Relevant or Not Relevant. If the expert cannot decide, he can mark the document as Skipped. It is recommended to minimize the number of Skip documents, as possible.

3. Once the expert has completed all 40 documents in the sample, the system calculates training status. Three states are possible – not stable, nearly stable or stable. Until stability is reached, the expert continues to the next sample. Stability usually requires 20-40 iterations, each of 40 documents.

4. The system reaches the Stabilization point when the marginal contribution of additional samples to the enhancement of the classifier approaches zero. Once it has been stable for five consecutive rounds, the system allows the user to calculate relevance scores for the entire population. Each document receives a relevance score in the range of 0 through 100.

5. During the training process, the system issues alerts if the expert provides inconsistent input. For example, if the expert marks one training document as not relevant, and a near-duplicate as relevant, the system will ask the expert to verify these tags.

CATCH-UP

1. The Relevance “catch-up” function manages incremental loads.
2. The aim of catch-up is to verify that the classifier built for the initial load is able to efficiently evaluate relevance for documents in the subsequent load. The catch-up phase comprises two phases:
   - Verification: when the new load is added, the system submits a sample of documents from the new load to the reviewer. If the new and previous loads
have the same characteristics (homogeneous), this means that the classifier will work well for the new load and the system will indicate that no additional training is required.

- Re-training: if the new and previous loads have different characteristics (heterogeneous), this means that the classifier for the previous load is not tuned for the new load, and re-training is required.

**DECISION**

1. The decision environment allows the user to select documents for inclusion in the review set. The user sets a cut-off relevance score. Documents with relevance scores below the cut-off mark are culled.

2. The Relevance decision-support environment provides information to help the user select the cut-off point. Based on the distribution of documents by relevance scores, the system generates review-relevance ratios. For example:
   - The user chooses a cut-off score that will require review 16% of the population
   - Based on user-entered review cost parameters, the system calculates the total review cost for reviewing 16% of the population. In this example, let's assume that the review cost for this volume of documents would be $266,000.
   - The system calculates that these documents (the top-scoring 16% of documents in the population) will yield 73% of the relevant documents (aka recall).
   - The user might decide to reduce risk by selecting a lower cut-off score. For example, the user might decide to review 30% of the collection. This cut-off policy will yield, say, 91% of the relevant documents. However, the total review cost will have increased to, say, $488,000.
   - Based on the proportionality considerations, the user will decide on the mix of risk and cost she is willing to bear, where risk is the percentage of relevant documents not included in the review set, and cost is the cost of review.

**VERIFICATION**

1. Perform quality assurance is to provide transparent verification of the results generated by the application.

2. Relevance supports a “Test the Rest” protocol to verify culling decisions. For example, using the distribution of relevance scores, the user may decide that documents with scores above 24 will be submitted for review. Documents with scores of 24 and below will be culled. This area of the collection is referred to as “The Rest” (aka the tail).
3. Let’s also assume that the decision support system shows that the Rest comprises 84% of the population, with recall of 12% (that is, 12% of the relevant documents) and precision of 0.5% (that is, 1 in every 200 documents is relevant). Test the Rest is intended to verify these statistics for the Rest.

4. Having defined the cut-off score (in our case 24), Test the Rest generates a random sample of documents from the Rest area.

5. The expert, acting as the Oracle, then reviews the sample documents, marking them as relevant or not. Using these tags, the system is able to make an independent calculation of recall and precision in the Rest zone. Equipped with this information, the user can confirm or modify the selected cutoff point. If hot documents are discovered in the Rest, or if there are significant discrepancies between the Test the Rest results and the results estimated in the decision environment, it may be necessary to revisit the assessment and training documents, or to resume training of the system.

6. The Test the Rest sample is designed to provide a confidence level of 95%. The default sample size is 500 documents. The margin of error depends on the percentage of relevant documents in the Rest zone. For example, if 5% of the Rest documents are found to be relevant, the margin of error is 1.9%. If 1% are relevant, the margin of error is 0.8%. The margin of error can be slightly reduced by taking a larger sample set.

7. Once verification is complete, engage with opposing counsel to communicate the metrics of the review set. If and where required, share relevant training documents, ensuring that privileged and non-relevant documents are not released. Reach agreement with opposing counsel regarding export format.