# **Ata de Reunião da NE-Soft-Factory**

Data 02/08/2018

Veículo: Skype

Início: 21:00 h Término: 22:21 h

Duração: 01:21 h

Participantes: Alfeu, Maria do Carmo, Helton, Fernando, Erick, Antônio Amorim, Antônio Filho

**Assuntos Tratados**:

* Fernando começou a reunião perguntando para Amorim sobre a revisão das heurísticas, Amorim falou que tinha encontrado alguns erros de gramática, mas que não corrigiu, Fernando pediu então para ele corrigir;
* Amorim falou que já está fazendo a revisão e Fernando falou que as heurísticas de Bolton precisam ser revisadas com mais atenção pelo fato de serem mais superficiais, mais fáceis, porém as respostas são muito parecidas, já as de Elsen têm mais detalhes;
* Antônio pesquisou o unidrive do Hotmail, ele promete trabalhar com OCR, mas, como os outros aplicativos, sem resolver as dificuldades;
* Fernando falou que não teve tempo para pesquisar o OCR, porque estava com as heurísticas;
* Alfeu disse que conversou com Helton e comentou que foi muito enfático com os riscos do OCR e deixou parecer que o problema era mais grave do que é na realidade;
* Fernando sugere criar uma lista no slack para todos postarem suas dificuldades;
* Alfeu falou que se poderia pensar em soluções alternativas, mas ainda é cedo para isto, ele descobriu um sistema que faz coisas parecidas com o nosso, o nome dele é “Querido Diário”, aberto à todos, incluindo o sistema buscador “Traz Diário”, ele disse que vai instalar e testar para ver se encontra pistas para solucionar os problemas do SISCAP;
* Alfeu perguntou se alguém havia pesquisado o OCR;
* Antônio falou que o problema não é assustador, mas merece atenção, o que ele mais se preocupa são as duas colunas e as tabelas, mas vai continuar as pesquisas. Ele falou que precisa e está pensando em um plano B, mas que depois que falou com Alfeu nasceu mais uma esperança;
* Alfeu falou que assim que viu o projeto “Querido Diário”, correu para falar com Antônio antes que ele falasse com o cliente auditor;
* Helton fez alguns testesno OCR do google e constatou que na maioria dos textos ele converte mais de 90% das palavras. O problema que ele viu foi que a divisão das colunas nas páginas, não é padronizada. Ele pede o compromisso da equipe para pesquisar e achar os problemas e as soluções. Ele pegou a página, identificou os elementos e fez a conversão, mas as tabelas quebraram;
* Antônio falou que as colunas são blocos de linhas como um parágrafo, se conseguisse ler a coluna como sendo uma linha, talvez fosse a solução. Ele obteve como resultado da conversão, um bloco com todos os nomes e outro com todas as matrículas, em um texto com duas colunas, nome e matrícula;
* Helton falou que pegou um texto com mais de duas colunas e que não conseguiu identificar o tipo de dado de cada coluna como no teste de Antônio;
* Alfeu falou que tem texto em que as linhas são mescladas, e que isso é um complicador a mais, que transformando em texto se perde a referência da organização da tabela. Falou também que as tabelas não têm valor importante para a auditoria, por isso poderia desconsiderá-las, mas tem que analisar antes, a importação dos dados das tabelas;
* Helton falou em fazer um tratamento só com as tabelas, para avaliar o que se pode fazer. Ele diz que os dados vêm misturados da conversão, observou que os outros sistemas similares vão em busca de textos dentro dos D.O., o que torna mais fácil o desenvolvimento do sistema, mas teria que mudar o projeto para fazer isto. Ele sugere que a equipe estude mais o projeto atual, uma semana, antes de tomar uma decisão de troca;
* Alfeu cita o exemplo de uma empresa que defendeu não utilizar o OCR evitando erros, ele diz que os erros mesmo num percentual menor, se forem recorrentes, serão malvistos pelo cliente;
* Helton falou que não dá para garantir percentuais de erros e acertos. Falou do status repórter e da avaliação independente, falou que se não conseguirmos resolver o problema do projeto, não irá chegar para o professor três meses depois e dizer que não deu certo e que também não tinha um plano B;
* Alfeu falou que dá tempo ainda para mudar o projeto, mas que não seria o ideal. Ele acha melhor adiar o início da sprint, porque ela pode mudar totalmente;
* A equipe concorda;
* Helton falou que deve haver uma data para apresentação do resultado dos estudos de OCR;
* Fernando defende a ideia de uma pessoa dar continuidade ao buscador e que as outras deveriam ficar com as pesquisas de OCR;
* Helton falou que poderia ser uma alternativa, mas poderia ser tempo perdido se não o OCR não desse certo;
* Fernando sugere então que a sprint 1 seja composta por duas frentes: uma para o buscador e outra para pesquisa do OCR;
* Alfeu falou que além de tudo isto, temos as nossas entregas aos professores que podem ficar comprometidas pela falta de tempo;
* Fernando falou em criar um quadro Burndown para a sprint;
* Alfeu falou que pode segurar o buscador por uma semana e aproveitar para focar no OCR, pelo risco que oferece. Ele falou em adiar o início da sprint para a próxima segunda-feira, todos ficariam pesquisando por sete dias e se der certo continua com o buscador depois. Ele falou que se deve resolver as incertezas do projeto;
* Helton lembrou que as entregas podem ser feitas, mesmo sem termos concluído a sprint;
* Alfeu falou que a meta é descobrir se o projeto pode dar certo ou não, e que vai passar o link do “Querido Diário” para todos testarem;
* Fernando quer que todos passem as suas pesquisas para a equipe, para evitar retrabalho;
* Alfeu falou que a melhor ferramenta até agora é a da google;
* Fernando vai procurar mais informações sobre esta ferramenta que Alfeu falou, no evento da google em que ele vai participar no final de semana;
* Alfeu falou que testou outras ferramentas indicadas por Fernando, mas que não foram tão boas quanto a da google. Ele falou que o foco é ver se a ferramenta da google é viável, senão, irá procurar por outra, e se não achar, irá pensar em mudar o projeto;
* Helton falou que vai disponibilizar o projeto do google no Git, para todos, e seguirá o tutorial da ferramenta. Ele disse que fez umas alterações e observou que a ferramenta trabalha com coordenadas x,y, e pede para o pessoal ver se isto seria uma facilidade;
* Antônio falou que encontrou ferramentas que transformam pdf em doc;
* Alfeu falou que fez estes testes com o adobe e conseguiu reconhecer os elementos do texto. Ele falou também, que o D.O. da União disponibiliza as publicações em html e xml, e que o sistema “Querido Diário”, é feito em python e que é idêntico ao nosso projeto;
* Amorim falou que está tudo muito subjetivo e que é melhor Antônio ver um plano B;
* Antônio achou um serviço cognitivo de imagem digitalizada da microsoft, concorrente ao google, que funciona similar ao google vision. Ele fez os testes e parece ser interessante aos nossos propósitos;
* Alfeu disse que a ideia é passar as alternativas encontradas para todos da equipe. Ele falou que está tentando falar com Felipe professor, para ver alguma equipe que já trabalhou com OCR e repassar a experiência para a nossa equipe. Alfeu falou que o atraso do buscador está sendo providencial para o adiantamento do OCR;
* Antônio falou que mandou o link da sua pesquisa, no slack;
* Alfeu falou sobre o calendário que criou no google e que o mesmo é colaborativo, todos podem alimentá-lo;
* Alfeu falou na divisão do trabalho, ele ficará estudando o “Querido Diário” e Helton, Fernando e Erick vão estudar a ferramenta do google utilizando várias fontes;
* Alfeu falou que o D.O. de São Paulo tem colunas também, e falou que depois avaliariam as ferramentas estudadas para ver qual a melhor para utilizar no projeto.