

Mestrado em Engenharia Informática

Dissertação/Estágio

Relatório Final

Reconhecimento de voz para interação com SI na área da saúde

Álvaro Miguel Granada Mateus

amateus@student.dei.uc.pt

Orientador do DEI:

Prof. Doutor Hugo Gonçalo Oliveira

Orientador da empresa:

Engenheiro Nuno Pereira

Data: 1 de Julho de 2016



FCTUC DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Mestrado em Engenharia Informática

Dissertação/Estágio

Relatório Final

Reconhecimento de voz para interação com SI na área da saúde

Álvaro Miguel Granada Mateus

amateus@student.dei.uc.pt

Jurís:

Prof. Doutor Pedro Furtado

Prof. Doutor Vasco Pereira

Data: 1 de Julho de 2016



FCTUC DEPARTAMENTO
DE ENGENHARIA INFORMÁTICA
FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE COIMBRA

Resumo

O presente estágio surge na certeza de que a comunicação verbal continua a ser o processo mais usual e natural para profissionais comunicarem entre si, e que encontrar uma solução capaz de intermediar essa interação é o ponto fulcral na procura por agilizar os processos existentes. É nesta premissa que, o presente estágio, procura por formas de interagir verbalmente com dispositivos móveis, nomeadamente *smartphones*, como forma de promover uma resposta rápida e automatizada às necessidades dos profissionais de saúde.

Assim sendo, o que se pretende com o presente estágio é avaliar e desenvolver uma arquitetura e prova de conceito de uma solução tecnológica capaz de fazer face as limitações existentes no processo comunicacional e, ainda, lidar com as dificuldades sentidas pelos profissionais de saúde no acesso e atualização de informação dos seus pacientes.

No final, a solução desenvolvida consistirá num agregado de serviços e lógica de processos com os quais se poderá integrar numa solução *mobile* e que permitirá disponibilizar para as unidades de saúde uma solução capaz de melhorar os mecanismos de comunicação existentes e, ainda, agilizar a execução dos processos internos com vista a melhorar os cuidados de saúde prestados.

Palavras-chave: eficiência, comunicação verbal, natural, engenharia de *software*

Agradecimentos

Na vida todos os momentos passam mas a experiência fica. Importa, contudo, lembrar das pessoas que fizeram parte dessa jornada que conseguiram acrescentar um sabor ainda mais doce.

Em primeiro lugar, gostaria de agradecer à minha família pelo esforço, dedicação e amor incondicional dados ao longo da minha vida de forma a que recebesse a educação e estaleca necessárias para fazer face aos desafios da vida. Fizeram de mim rapaz e homem.

Ao Professor Doutor Hugo Oliveira, agradeço a orientação e disponibilidade dada de forma a fazer face a todos os desafios existentes no presente estágio. Deixo, ainda, uma nota adicional de agradecimento ao Professor Doutor Fernando Perdigão e ao aluno de Doutoramento Jorge Proença do Departamento de Engenharia Eletrotécnica da Universidade de Coimbra pela ajuda fornecida na compreensão e especificação de testes às ferramentas de reconhecimento de voz.

Aos meus amigos tenho a agradecer tudo. Apesar da influência clara dos meus pais na minha educação e nos valores transmitidos, tenho de realçar também a importância de todos os meus amigos na definição da minha pessoa. Ainda que nem todos estejam constantemente presentes todos fazem parte de mim nos momentos partilhados ao longo do tempo. Sejam eles momentos de preocupação, diversão ou de apoio. No entanto, será por esta montanha russa de experiências que eu continuarei sempre à procura, todos os dias, por novas experiências e momentos uma vez que, no final, seja a chorar, sorrir ou uma mistura dos dois, é na companhia deles que de certeza irei encontrar mais momentos de felicidade.

Em último lugar, mas de todo menos importante, gostaria de expressar o meu obrigado à MedicineOne pela oportunidade dada no desenvolvimento deste projeto mas, acima de tudo, pelo respeito, apoio, carinho e humor com que me receberam e trataram ao longo deste período. Criei aqui amizades e respeito por toda uma cultura de interação de igualdade em ambiente de trabalho. Apesar de as condições atuais não o permitirem, sei que a vida é preenchida de oportunidades e, por isso, pode ser que no futuro se proporcione um regresso a esta família. Deixo uma palavra de força e desejos de boa sorte para um futuro risonho e próspero.

Lista de Acrónimos

AMR – *Adaptative Multi-Rate*

API – *Application Programming Interface*

ASR – *Automatic Speech Recognition*

CPLP – *Comunidade dos Países de Língua Portuguesa*

DI – *Dependency Injection*

EDPD – *European Data Protection Directive*

EHR – *Electronical Health Record*

EVRC – *Enhanced Variable Rate Codec*

FLAC – *Free Lossless Audio Codec*

HIPAA – *Healthcare Insurance Portability and Accountability Act*

HTTP – *Hypertext Transfer Protocol*

IoC – *Inversion of Control*

ISP – *Interface Segregation Principle*

LPCM – *Linear-16*

LSP – *Liskov Substitution Principle*

MEF – *Managed Extensibility Framework*

ML – *Machine Learning*

MPEG – *Moving Picture Experts Group*

MP3 – *MPEG Layer 3*

NLU – *Natural Language Understanding*

OCP – *Open/Closed Principle*

OS – *Operating System / Sistema Operativo*

PCM – *Pulse Code Modulation*

PCMU – *Pulse Module Code mu-law*

QCELP – *Qualcomm code-excited linear prediction*

SaaS – *Software-as-a-Service*

SDK – *Software Development Kit*

SOA – *Service-Oriented Architecture*

SRP – *Single Responsibility Principle*

STT – *Speech to Text*

Lista de Figuras

Figura 1 - Ilustração do processo de alta de um utente numa unidade hospitalar	4
Figura 2 - Planeamento das tarefas referentes ao primeiro semestre	4
Figura 3 - Ilustração do plano de trabalhos seguido	5
Figura 4 - Planeamento das tarefas referentes ao segundo semestre	6
Figura 5 - Composição de um sistema conversacional, imagem adaptada de (Jurafsky & Martin, 2008)	12
Figura 6 - Assistentes Pessoais dos principais sistemas operativos móveis - Google Now, Siri e Cortana	16
Figura 7 - Assistente Pessoal da api.ai - Assistant.ai	17
Figura 8 - Dispositivo Google Assistant apresentado no Google I/O 2016 no contexto da domótica	17
Figura 9 - Exemplo de reconhecimento de ações e entidades da plataforma wit.ai	29
Figura 10 - Visão global da arquitetura do sistema Voccallia	51
Figura 11 - Vista em Camadas do Voccallia	54
Figura 12 - Perspetiva dinâmica da arquitetura - Descoberta da intenção do utilizador a partir do áudio associado ao pedido do utilizador	55
Figura 13 - Descrição dos contratos dos serviços disponibilizados para os clientes do Voccallia	57
Figura 14 - Interoperabilidade entre diferentes sistemas de clientes móveis e o Voccallia	58
Figura 15 - Representação dos contratos de comunicação entre o conetor do EHR e o Voccallia ..	59
Figura 16 - Mecanismo de interoperabilidade entre o Voccallia e diferentes EHRs - EHR address lookup	61
Figura 17 - Estratégia de delegação da autenticação do utilizador usada no Voccallia - Brokered authentication	61
Figura 18 - Solução arquitetural de plugins para o módulo VoccalliaCognitive	63
Figura 19 - Solução arquitetural de plugins para o módulo VoccalliaSenses	64
Figura 20 - Design Pattern de Repositório usado para abstrair o acesso aos dados, imagem (The Repository Pattern, 2016)	65
Figura 21 - Solução arquitetural definida para o módulo VoccalliaOrchestration	66
Figura 22 - Modelo conceptual de dados do Voccallia	68
Figura 23 - Composição e descoberta de dependências com o MEF, imagem adaptada (htt)	69

Figura 24 - Domínio de integração do conector desenvolvido	72
Figura 25 - Principais menus da aplicação desenvolvida	73
Figura 26 - Plataforma web desenvolvida para a recolha de amostras de áudio	77
Figura 27 - Modo de geração aleatória	78
Figura 28 - Modo de geração normal	78
Figura 29 - Resultado obtido do HTK do processo comparação de duas transcrições (original e a obtida do ASR)	81
Figura 30 - Comparativo da percentagem média de WER por categoria de cenário de utilização ..	83
Figura 31 - Comparativo da percentagem média de WER para os dois sexos (Masculino/Feminino)	83
Figura 32 - Análise de resultados da percentagem média de WER por utilizador e género	85
Figura 33 - Avaliação global do WER nos dados de teste para cada uma das ferramentas de ASR (Google e Nuance	86

Lista de Tabelas

Tabela 1 - Identificação dos riscos associados à realização do projeto durante o período de estágio	1
Tabela 2 - Planos de mitigação dos riscos identificados com maior impacto na realização do projeto	2
Tabela 3 - APIs e SDKs suportados pela plataforma api.ai	31
Tabela 4 - Formatos de áudio e taxas de amostragem suportadas por cada um dos motores de ASR	32
Tabela 5 - Comparativo das ferramentas conforme os critérios definidos para o contexto da solução	33
Tabela 6- Identificação e descrição dos atores intervenientes no sistema	36
Tabela 7 - Decomposição do ator Profissional de Saúde em outros atores	36
Tabela 8 - Definição das classes de priorização dos requisitos funcionais	38
Tabela 9 - Resumo dos requisitos funcionais levantados para o projeto	39
Tabela 10 - Cenário concreto de execução do requisito de interoperabilidade 1	41
Tabela 11 - Cenário concreto de execução do requisito de interoperabilidade 2	42
Tabela 12 - Cenário concreto de execução do requisito de interoperabilidade 3	42
Tabela 13 - Cenário concreto de execução do requisito de interoperabilidade 4	43
Tabela 14 - Cenário concreto de execução do requisito de interoperabilidade 5	43
Tabela 15 - Cenário concreto de execução do requisito de extensibilidade	44
Tabela 16 - Cenário concreto de execução do requisito de modificabilidade 1	45
Tabela 17 - Cenário concreto de execução do requisito de modificabilidade 2	45
Tabela 18- Stakeholders e necessidades de documentação para o projeto Voccallia	49
Tabela 19 - Stakeholders e necessidades de documentação específicas para o âmbito do estágio	50

Conteúdo

1	Introdução.....	1
1.1	Motivação.....	1
1.2	Projeto Voccallia.....	5
1.3	Objetivos	5
1.4	Estrutura do Documento.....	6
2	Metodologia.....	9
2.1	Metodologia de desenvolvimento	9
2.2	Processos de Engenharia.....	10
2.2.1	Estimativas	10
2.2.2	Gestão de Riscos.....	10
2.2.3	Planeamento e Tracking.....	3
2.2.4	Monitorização	7
3	Estado da Arte.....	11
3.1	Abordagens na Literatura.....	11
3.2	Trabalho Relacionado.....	14
3.2.1	Assistentes Pessoais: Google Now, Siri, Cortana, Assistant.ai e Personal Medical Advisor	15
3.2.2	Sistemas de documentação: Nuance, Winscribe e VoiceInteraction	18
3.2.3	Sistema de integração: <i>Vocera</i> , Nuance	19
3.2.4	Discussão	21
3.3	Análise de ferramentas	22
3.4	Discussão.....	32
4	Análise e Especificação de Requisitos.....	35
4.1	Atores do Sistema	36
4.2	Requisitos Funcionais.....	37
4.3	Requisitos Não-funcionais.....	40
4.3.1	Interoperabilidade.....	40
4.3.2	Modificabilidade.....	44
4.3.3	Testabilidade	46

5	Análise e Desenho da Arquitetura	47
5.1	Identificação das Necessidades de Documentação	48
5.2	Descrição da Visão Global do Sistema	50
5.3	Decisões Arquiteturais	52
5.4	Arquitetura desenvolvida.....	52
5.4.1	Estrutura e Organização	53
5.4.2	Análise das propriedades sistémicas e <i>tradeoffs</i>	56
6	Trabalho Desenvolvido	67
6.1	Desenvolvimento da Solução	67
6.2	Desenvolvimento dos Sistemas de Integração	71
6.2.1	Desenvolvimento do Conetor <i>MedicineOne</i>	71
6.2.2	Desenvolvimento da aplicação móvel.....	72
7	Testes	75
7.1	Testes de aceitação	75
7.2	Testes de Integração	75
7.3	Plataforma de recolha de amostras de áudio para teste.....	76
7.4	Testes às ferramentas de ASR	80
8	Considerações Finais.....	87
8.1	Ponto de situação.....	89
8.2	Principais obstáculos	89
8.3	Trabalho futuro	90

1 Introdução

O presente documento tem como objetivo apresentar o trabalho relativo à proposta de estágio de criação de um sistema de “Reconhecimento de voz para interação com os sistemas de informação (SI) na área da saúde”, desenvolvido pelo aluno Álvaro Miguel Granada Mateus no âmbito da disciplina de “Dissertação/Estágio” referente ao Mestrado de Engenharia Informática, ramo de Engenharia de *Software*, da Universidade de Coimbra, no ano letivo de 2015/2016.

A realização do mesmo decorreu na empresa MedicineOne, sediada no Instituto Pedro Nunes (IPN), sob a orientação do Professor Doutor Hugo Gonçalo Oliveira do Departamento de Engenharia Informática (DEI) e do Engenheiro Nuno Pereira da MedicineOne.

Neste capítulo é apresentada uma contextualização do projeto a desenvolver, as suas motivações e objetivos a cumprir e, também, das contribuições esperadas da parte do aluno. Por fim, é apresentada uma visão global da estrutura e conteúdos do documento.

1.1 Motivação

Nos dias de hoje, cada vez mais somos “disciplinados” para a realização de técnicas de trabalho mais eficientes. De fato, sejam engenheiros civis, informáticos, eletrotécnicos, médicos, enfermeiros, terapeutas e, até mesmo, desportistas de alta competição há algo que todas estas profissões partilham que é a clara tendência pela procura e necessidade da prática de mecanismos de melhoria da eficiência. Desta forma, apesar da clara diferença de contextos, a procura e identificação das fraquezas nos processos de cada um permite que sejam realizados estudos e testes com o objetivo de se alcançarem melhorias.

Deste modo, enquadrando a análise para o domínio da saúde, podemos realçar o papel que a comunicação interpessoal assume no funcionamento das infraestruturas de saúde. Na verdade, a garantia da realização de uma comunicação eficiente entre os diferentes profissionais deste tipo de infraestruturas é implícita dada a **imprevisibilidade** característica destes ambientes, associada ao ter de prestar cuidados de saúde aos pacientes na forma e tempo corretos. Tal acontece porque a prestação dos cuidados de saúde pressupõe a coordenação de diversas equipas multidisciplinares que colaboram entre si com vista a completarem diferentes fases de um processo de cuidados.

Note-se que é neste tipo de ambientes, onde a **exigência de uma comunicação rápida e eficaz**, é indubitavelmente **crítica** pelas seguintes razões:

- De forma a assegurar o estado de saúde dos pacientes numa unidade, por vezes pode ser necessária a tomada de decisões e ação num curto período de tempo;
- Para a realização de um ato médico, pode ser necessária a intervenção de profissionais de saúde de diferentes áreas;
- Devido à grande magnitude deste tipo de infraestruturas, pode haver uma grande dispersão dos profissionais pelas diferentes unidades, complicando assim a sua localização
- Por outro lado, a localização física dos profissionais é um desafio acrescido devido à grande mobilidade dos mesmos na prestação dos cuidados de saúde;

Efetivamente, face aos pontos identificados, **as soluções tradicionais de comunicação** existentes neste tipo de infraestruturas (ex. telefones fixos para comunicação entre profissionais) **não são suficientes nem adequadas** para as necessidades existentes. Isto acontece porque de forma a efetuar a localização de um profissional é necessário saber onde o mesmo se encontra a cada momento algo que, com este tipo de solução, é impraticável dada a grande mobilidade que pode existir entre as unidades de serviço. Contudo, **o uso de telemóveis pessoais** permite o acrescento de múltiplas vantagens (ex. lidar com o problema da mobilidade ao contactar o profissional diretamente em vez da unidade de serviço) ainda que, na ausência de um meio complementar de gestão da informação, seja também uma **solução inviável e incapaz** de colmatar as necessidades identificadas para o presente contexto pelas seguintes razões:

- A realização da comunicação entre profissionais pressupõe a necessidade de se conhecer quais os profissionais que se encontram de serviço no momento;
- Não existe um mecanismo central de disponibilização e gestão da informação atual de presença dos profissionais;
- A inserção e atualização de informação de utentes continua a exigir a necessidade de deslocação até a um computador;
- Nem todos os profissionais têm os contactos pessoais dos restantes;
- A comunicação realizada entre dispositivos pessoais não é registada em nenhum lugar (perde-se o registo temporal e espacial da informação);

Na verdade, e apesar de, na sua generalidade, as unidades de saúde funcionarem em condições satisfatórias, nem sempre os processos utilizados são os mais eficientes. No contexto hospitalar, por

exemplo, os casos relacionados com a admissão, transferência/transporte e alta de pacientes são aqueles mais perceptíveis em que a comunicação entre diferentes pessoas e unidades de serviço de uma unidade hospitalar é menos eficiente. Tal, deve-se ao fato de poderem ser necessárias diferentes pessoas e ações para se conseguir alcançar um único fim, como se pode verificar da análise do processo de alta de um paciente, ilustrado na Figura 1.

Como se pode constatar, o processo de alta do doente é complexo e envolve, na sua maioria, várias relações de interdependência entre as suas diferentes etapas. Desta forma, a rapidez com que o processo de alta de um doente ocorre está intrinsecamente relacionada com a rapidez com que a comunicação entre os diferentes profissionais é realizada. Isto porque durante todo o processo de comunicação verbal e digital entre os diferentes intervenientes (ex. médico e enfermeiro ou enfermeiro e auxiliares de saúde) não existe a garantia de a comunicação ser realizada em instantes consecutivos e de forma ordenada por paciente, tendo com resultado o paciente ter que esperar por longos períodos até que o processo de alta do mesmo seja retomado pela equipa de enfermagem ou pelo médico responsável.

Note-se que existem múltiplos fatores que contribuem para a ocorrência deste tipo de cenários, como é o caso de situações de sobreposição na troca de turnos da equipa de enfermagem e médica e a ordem médica não ser transmitida; situações nas quais os médicos responsáveis por uma unidade de serviço não se encontram na unidade para apreciação do doente (ex. prestação de cuidados noutras unidades de serviço); ou situações onde outros pacientes assumem um estatuto de prioridade de prestação dos cuidados de saúde superior.

De fato, tal é explicado pela comunicação entre profissionais atualmente ser realizada, quase exclusivamente, de forma verbal e com a necessidade de interação manual com o sistema de gestão de informação da unidade de saúde (normalmente, o enfermeiro ou médico aproveita numa fase posterior à avaliação dos pacientes para interagir com o sistema ou com os enfermeiros de serviço de forma a desencadear as restantes fases do processo) o que torna todo o processo em geral bastante ineficiente e moroso.

E é neste contexto que iremos apresentar o projeto *Voccallia* na secção seguinte.

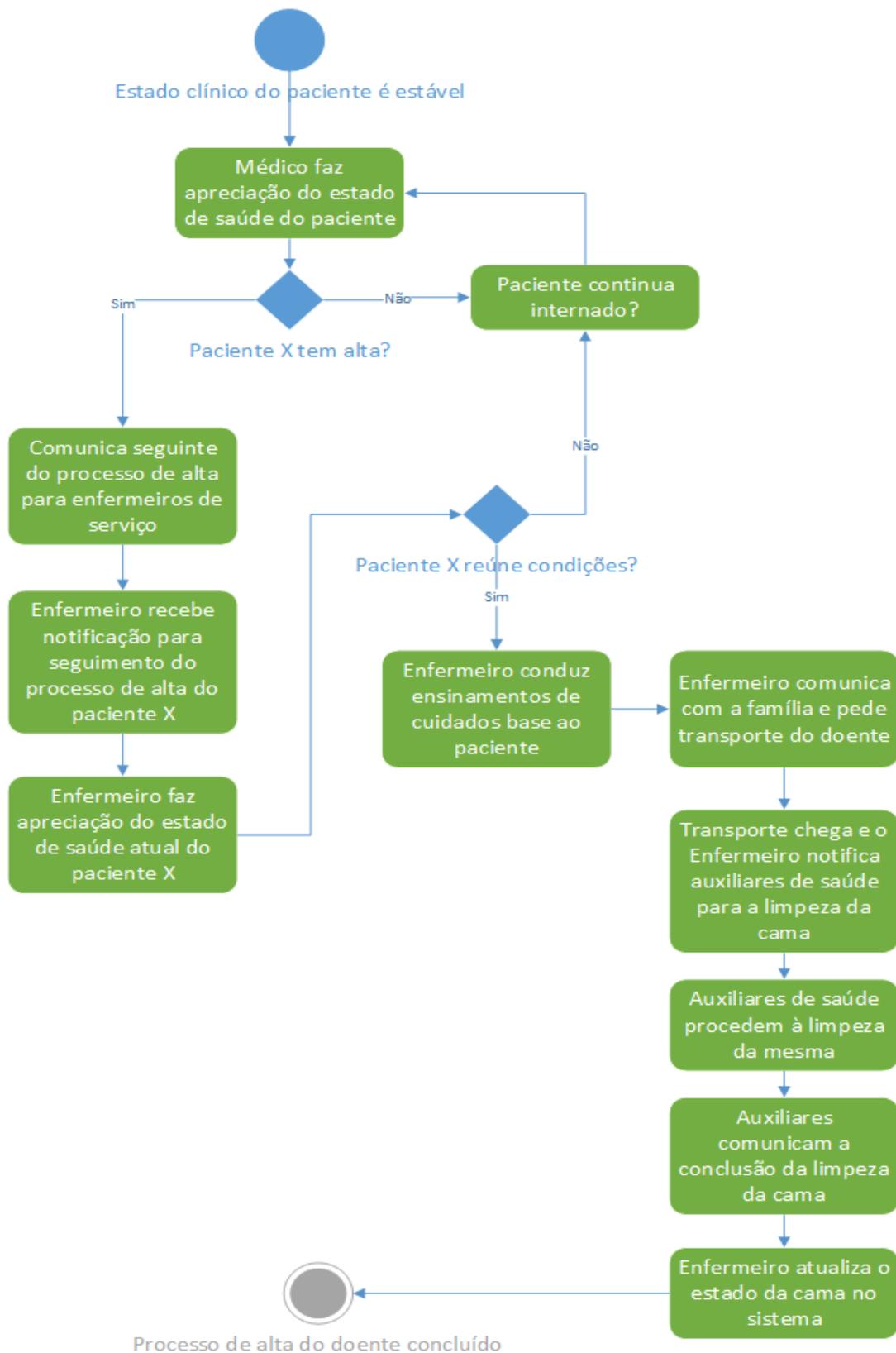


Figura 1 - Ilustração do processo de alta de um utente numa unidade hospitalar

1.2 Projeto Voccallia

O projeto *Voccallia* surge, com o presente estágio, no sentido de fazer face aos problemas identificados com base na premissa de que a comunicação verbal continua a ser o meio mais eficiente de comunicação. Assim, e tendo em conta os avanços significativos nas tecnologias de reconhecimento de voz, em inglês *Automatic Speech Recognition (ASR)*, e compreensão de linguagem natural, em inglês *Natural Language Understanding (NLU)*, o objetivo deste projeto é o de avaliar a possibilidade de criação de uma solução, baseada nestas tecnologias, a fim de melhorar e agilizar os processos de comunicação entre profissionais e, conseqüentemente, melhorar e integrar os diferentes processos de gestão e de ação característicos das empresas.

Desta forma, e atuando a MedicineOne na área da saúde, o principal foco deste projeto é naturalmente dirigido a unidades de saúde, mais especificamente, unidades de hospitalares e unidades de cuidados continuados. Com efeito, dadas as suas necessidades são estas as que mais beneficiarão de um sistema deste tipo. No entanto, como já fora referido, muitas outras áreas de negócio poderão também usufruir de uma solução semelhante.

1.3 Objetivos

O presente estágio prevê o desenvolvimento de uma prova de conceito de uma arquitetura de serviços em torno das funcionalidades de ASR e de NLU no contexto da saúde.

Desta forma, propôs-se que o estagiário teria os seguintes objetivos globais:

- Projetar e desenvolver uma arquitetura de *software* que garanta o cumprimento dos seguintes requisitos de qualidade:
 - Interoperabilidade – aplicações e serviços existentes no domínio MedicineOne e outros Sistemas de Informação de Saúde, em inglês *Electronical Health Record (EHR)* devem conseguir trocar informações significativas entre si e de forma útil;
 - Modificabilidade – o sistema deve ser capaz de acomodar alterações em componentes específicas com o intuito de refinar e/ou corrigir eventuais erros e/ou falhas presentes em módulos, serviços e/ou processos;
 - Testabilidade – o desenvolvimento dos módulos, serviços e processos deve ser realizado de forma a garantir a sua robustez e validade.
- Aprendizagem e aplicação das tecnologias (ex. C#, *Windows Communication Foundation - WCF*) usadas pela MedicineOne para o desenvolvimento da solução proposta.

- Definição e aplicação de estratégias de documentação adequadas aos objetivos anteriormente referidos.

Por outro lado, os principais objetivos a alcançar são os seguintes:

- Criação de serviços para interação com clientes móveis a fim de expor as seguintes funcionalidades:
 - Envio de comandos pelo utilizador em formato voz (suporte áudio) para interpretação da intenção do utilizador e reconhecimento das entidades envolvidas;
 - Envio de comandos pelo utilizador no formato texto para interpretação da intenção do utilizador e reconhecimento das entidades envolvidas;
 - Execução de ações, previamente definidas e reconhecidas pelo sistema e pelo utilizador, para o EHR;
- Desenvolvimento das componentes que disponibilizam as funcionalidades de ASR e NLU (dependências externas) utilizadas pelos serviços expostos ao utilizador.
- Desenvolvimento de um mecanismo de tradução de intenções (ações a realizar) do utilizador em chamadas a serviços de um EHR que, por norma, são representativos de processos de negócio do próprio. Por outro lado, podem também corresponder a mecanismos de comunicação interna entre utilizadores.
- Desenvolvimento de uma infraestrutura de serviços de suporte às ações disponíveis ao utilizador e que intermedie e abstraia a comunicação entre os serviços previamente definidos e os serviços existentes no EHR ao qual o utilizador se encontre associado (ex. MedicineOne).

1.4 Estrutura do Documento

O presente documento encontra-se estruturado para que o leitor entenda, em primeiro lugar, qual o ambiente e as condições (definição do problema, motivação e objetivos) que proporcionaram a iniciativa de desenvolvimento deste projeto e, posteriormente, seja orientado progressivamente para os processos, funcionalidades e requisitos do produto a desenvolver durante o projeto, tendo sempre como base de trabalho a análise e avaliação de todas as ferramentas consideradas. Desta forma, o documento encontra-se organizado de acordo com a seguinte estrutura:

- O primeiro capítulo introduz o leitor para o ambiente e os propósitos que caracterizam a realidade da empresa MedicineOne e para as necessidades dos seus clientes.
- O capítulo 2 expõe quais as metodologias adotadas no desenvolvimento do projeto. Ou seja, neste capítulo é definida a organização de tarefas no âmbito do projeto, desde o desenvolvimento da camada de aplicação ao desenvolvimento dos serviços que usam o motor semântico, e as metodologias próprias de gestão e planeamento de projetos – ciclo de vida adotado, métodos de planeamento, estimação e acompanhamento de tarefas, gestão de riscos, planeamento e *tracking*, e monitorização.
- O terceiro capítulo apresenta e categoriza o trabalho relacionado existente nas diversas áreas de negócio. Aqui são descritas algumas soluções e como cada uma usa as tecnologias de ASR e NLU. Posteriormente, são descritas algumas ferramentas destas tecnologias para a resolução do problema identificado no primeiro capítulo. No fim, é feito um resumo das tecnologias estudadas e apresentadas as conclusões retiradas da análise realizada com vista a identificação das ferramentas a usar na implementação da solução do presente estágio.
- O quarto capítulo descreve os requisitos funcionais e não funcionais do sistema a implementar através da representação dos principais atores do sistema, de cenários concretos de, descrição do processo e especificação das táticas arquiteturais a usar.
- O quinto capítulo descreve a arquitetura idealizada para fazer face aos requisitos especificados no quarto capítulo. Aqui são identificadas as principais necessidades de documentação da arquitetura, descrição da sua visão global, decisões arquiteturais tomadas e, por fim, analisada a sua estrutura e principais propriedades sistémicas.
- O sexto capítulo apresenta o trabalho desenvolvido durante o presente estágio.
- O sétimo capítulo descreve os testes realizados e ferramentas criadas para o suporte à realização dos mesmos. No final, são apresentados os testes realizados no âmbito das ferramentas de ASR e a análise dos seus resultados.
- O oitavo e último capítulo apresenta as considerações finais do estagiário relativamente ao trabalho desenvolvido no projeto. Aqui é apresentado um ponto de situação do projeto, principais obstáculos presentes durante a sua realização e trabalho futuro.

2 Metodologia

O presente estágio prevê que, dado o caráter introdutório do projeto e as diferentes fases que caracterizam o processo de estágio, diferentes processos de desenvolvimento e organização sejam utilizados com vista a beneficiar uma correta concretização dos objetivos do projeto. Desta forma, o estagiário beneficiará do acompanhamento da parte de diferentes elementos da empresa para as diferentes fases envolvidas (ex. especificação e análise de requisitos - Arquiteto e Cliente; Especificação da Arquitetura - Arquiteto e Equipa de desenvolvimento móvel; Prototipagem - Cliente, Equipa de desenvolvimento móvel).

Assim, as próximas secções do presente capítulo têm como objetivo o de explicitar os diferentes processos de engenharia e metodologias de desenvolvimento utilizados.

2.1 Metodologia de desenvolvimento

A metodologia de desenvolvimento de um projeto deve ser devidamente ponderada e avaliada de acordo com as necessidades e características de projeto. Efetivamente, a escolha por uma metodologia nem sempre é uma tarefa trivial uma vez que o mundo e as pessoas que nos rodeiam não se mantêm constantes. Existe uma constante alteração de ambientes e pessoas. Da mesma forma, um projeto de *software* está sujeito às mesmas condições. Se por um lado os requisitos podem não se manter iguais ao longo do tempo já que as necessidades atuais requerem uma constante adaptabilidade dos sistemas, por outro lado, nem sempre se consegue lidar com essa adaptabilidade tendo em conta todas as variáveis (tempo necessário, orçamento, viabilidade, etc.) apesar de se poderem usar estratégias para minimização das consequências adjacentes.

Deste modo, e tendo em conta que o presente projeto está sujeito a uma contínua avaliação das funcionalidades e do comportamento delas esperado, a metodologia que se definiu para a realização do projeto foi uma **metodologia ágil** baseada num **processo incremental e centrado no desenvolvimento de *features* (FDD)** (Hunt, 2006). Assim, pode-se modularizar todo desenvolvimento dos serviços em torno das funcionalidades e casos de uso característicos da aplicação e do motor semântico. Posteriormente, podem ser integrados com o protótipo desenvolvido pela equipa de desenvolvimento móvel e, assim, ao manter-se o desenvolvimento dividido e organizado em etapas, toda a documentação, testes e certificação de qualidade será realizada e mantida conforme a finalização dos módulos e sua integração. Adicionalmente, sendo

este um projeto **intimamente ligado às ações dos utilizadores alvo**, esta metodologia permite manter um contato mais frequente com o cliente, assegurando uma **maior facilidade na sua gestão e possíveis alterações aos requisitos**.

2.2 Processos de Engenharia

Nesta secção são apresentados os processos de engenharia que serão tidos em conta no período de desenvolvimento referente ao estágio.

2.2.1 Estimativas

Para efeitos do estágio utilizou-se a técnica de **Three-point estimation**¹ como mecanismo de estimação do tempo necessário para realizar as tarefas. Note-se que, para além do fato de haver uma forte componente de investigação associada ao estudo de ferramentas de processamento semântico e de reconhecimento de voz, há ainda desafios de desenvolvimento e validação dos serviços envolvidos nestas ferramentas. Adicionalmente, para efeitos de facilidade de integração dos próprios serviços com a arquitetura de serviços já existente da empresa, está definido que o desenvolvimento dos mesmos seja realizado em tecnologias *.NET*. Dada a inexperiência do estagiário no uso das tecnologias, definiu-se que estimação do tempo necessário para a realização de cada tarefa seja efetuado sempre tendo em conta três medidas: **Most likely case**, **Best case** e **Worst case**.

No final, o tempo de cada tarefa é calculado a partir dos três de acordo com a seguinte fórmula:

$$(Best\ Case + (4 * Most\ Like\ case) + Worst\ case) / 6$$

2.2.2 Gestão de Riscos

A ideia de que os projetos de Eng^a de Software decorrem sempre como previsto é, claramente, uma ilusão. Seja porque os requisitos foram mal elaborados e/ou compreendidos ou por falta de experiência, técnica e recursos ou, ainda, por imprevistos que levam ao atraso do desenvolvimento das funcionalidades. Ou seja, todos os projetos de *software* estão suscetíveis à existência de um conjunto de riscos durante o seu desenvolvimento. A forma como se consegue realizar a sua identificação e gestão é que faz a diferença na capacidade de alcançar o sucesso.

¹ https://onlinecampus.bu.edu/bbcswebdav/pid-904282-dt-content-rid-2557540_1/courses/13sum1metad644sol/lecture02a/metad644_L02T04d_threepointestimates.htm

Para este efeito, é necessária uma clara identificação de quais os riscos reais associados ao projeto e as suas potenciais consequências. Por outro lado, interessa também saber qual a probabilidade associada ao risco bem como em que altura ele tem impacto no projeto. Esta análise é realizada na subsecção.

Por fim, realizada a análise dos riscos interessa desenvolver um conjunto de estratégias capazes de minimizar eventuais ocorrências dos mesmos. A sua análise é realizada na subsecção 2.2.2.1 - Identificação de Riscos.

2.2.2.1 Identificação de Riscos

A presente subsecção tem como objetivo a identificação clara dos riscos associados ao projeto e qual o impacto de cada um quanto à possibilidade de atraso no plano de trabalhos ou, ainda, da eventualidade de **impedir a realização ou sucesso do projeto na sua totalidade**. Desta forma, na Tabela 1 é apresentado quais os riscos que se identificaram e a janela de tempo para a qual se espera a sua ocorrência durante o período do presente trabalho.

Assim, para o período mencionado importa, ainda, categorizar a probabilidade associada à ocorrência de cada um destes riscos e o impacto associado:

- **Probabilidade de ocorrência:**
 - **Muito Provável** – de 75 a 100 %
 - **Provável** – de 50 a 74 %
 - **Pouco Provável** – abaixo de 50 %
- **Impacto no projeto**
 - **Crítico** – compromete o sucesso do projeto
 - **Elevado** – consegue-se atingir os objetivos definidos com esforço adicional
 - **Médio** – consegue-se atingir os objetivos definidos dentro do esforço previsto

As definições descritas são utilizadas na Tabela 2 para ajudar a definir os planos de mitigação para os riscos a colmatar de acordo com a estratégia apresentada na secção Mitigação de Riscos.

ID do Risco	Fato	Consequência	Prazo
1	O idioma português (PT-PT) ainda é pouco suportado por ferramentas de reconhecimento de voz.	Possibilidade de construção de uma solução de reconhecimento de voz <i>in-house</i> com o intuito de suportar o idioma.	Curto
2	Ferramentas de reconhecimento de voz ainda não são perfeitas – podem falhar no reconhecimento de algumas palavras	Falha no reconhecimento de entidades e/ou ações a realizar no motor semântico.	Longo
3	Existem muitas ferramentas de detecção e reconhecimento de voz	Avaliação de cada uma das ferramentas torna-se um processo incomportável para o tempo disponível.	Curto
4	Existem muitas ferramentas de processamento e interpretação de linguagem natural	Avaliação de cada uma das ferramentas torna-se um processo incomportável para o tempo disponível.	Curto
5	Diferentes utilizadores têm diferentes formas de se exprimir para um mesmo conjunto de ações em linguagem natural.	Aumento da complexidade na especificação do motor semântico capaz de reconhecer as diferentes formas de expressão.	Médio
6	Diferentes utilizadores têm diferentes acentuações nas palavras e sotaques.	Maior probabilidade de erro no reconhecimento de entidades e ações a partir da voz.	Médio
7	Limitações de <i>hardware</i> existentes nos dispositivos dos utilizadores nas tarefas de captura do áudio-	Qualidade do áudio resultante da captura de voz pode não ser suficientemente boa para permitir o correto reconhecimento das palavras.	Longo
8	O acesso a algumas ferramentas de ASR é condicionado (ex. aprovação de registo de projeto, quotas de uso).	A realização dos testes de comparativos com outras ferramentas é comprometido, podendo não ser apresentados resultados de análise.	Curto
9	O acesso a algumas ferramentas de NLU é condicionado (ex. aprovação de registo de projeto, quotas de uso).	A avaliação da ferramenta como potencial alternativa fica condicionada pela falta de testes comparativos e de integração.	Curto

Tabela 1 - Identificação dos riscos associados à realização do projeto durante o período de estágio

2.2.2.2 Mitigação de Riscos

Esta subsecção tem como intuito ajudar a perceber quais as estratégias utilizadas para a mitigar a existências dos riscos identificados na subsecção anterior. Uma vez que ainda foram identificados um bom número de riscos associados ao projeto, optou-se por se realizar uma estratégia de mitigação conhecida por **Pareto Top N²** a partir da qual se define a mitigação de um conjunto restrito de riscos de acordo com os níveis de interesse (**Impacto, Probabilidade e Janela de tempo**). Desta forma, optou-se por se **mitigar os riscos que têm influência a curto e médio prazo, com uma probabilidade de ocorrência acima dos 50%** e que a sua realização compromete o sucesso do projeto. A Tabela 2 apresenta os planos de mitigação para os sete riscos mais significativos do projeto.

ID do Risco	Nível de probabilidade	Impacto no projeto	Mitigação
1	Muito Provável	Crítico	Investigação do nível de suporte ao idioma pelas principais empresas na área.
2	Muito Provável	Crítico	Mecanismos de confirmação de ações e entidades presentes no discurso realizado.
3	Provável	Elevado	Restrição da análise de ferramentas com base num conjunto de critérios (secção 3.3).
4	Provável	Elevado	Restrição da análise de ferramentas com base num conjunto de critérios (secção 3.3).
5	Provável	Crítico	Restrição das formas de expressão com o motor semântico para um conjunto pré-definido.
8	Provável	Elevado	O desenvolvimento e teste a partir destas ferramentas aquando da sua análise é adiado para uma fase futura de prospeção de mercado.
9	Provável	Elevado	O desenvolvimento e teste a partir destas ferramentas aquando da sua análise é adiado para uma fase futura de prospeção de mercado.

Tabela 2 - Planos de mitigação dos riscos identificados com maior impacto na realização do projeto

² http://sce.uhcl.edu/helm/BB-TestRiskMan/my_files/module6/topn.htm

2.2.3 Planeamento e Tracking

As tarefas de planeamento e de *tracking* permitem, se realizadas corretamente, uma organização clara e eficiente de todas as tarefas envolvidas na realização do projeto.

Portanto, para este projeto optou-se por se usar o **diagrama de Gantt** uma vez que este permite uma clara hierarquização das tarefas de acordo com o seu nível de prioridade, a definição do tempo gasto para a finalização de cada tarefa e, ainda, estabelecer relações de precedência entre as mesmas. Consequentemente, torna-se possível efetuar diferentes análises, sendo que a que mais se destaca é a visualização de eventuais demoras na execução de determinadas tarefas que causem atrasos no desenvolvimento do projeto.

Por outro lado, a própria definição de relações de precedência faz com que o desenvolvimento seja mais estruturado pelo simples fato de ser evidente quais as tarefas que devem ser executadas em primeiro lugar com vista a determinado objetivo ser alcançado.

A Figura 2 apresenta o planeamento das tarefas realizadas durante o primeiro semestre de onde é perceptível a clara aposta nas fases de estudo de soluções e tecnologias já existentes em detrimento de uma análise de requisitos e desenho da arquitetura dada a natureza exploratória do presente estágio.

Por outro lado, a Figura 3 e Figura 4 representam os planos de trabalhos seguidos e idealizados para as tarefas desenvolvidas no **segundo semestre**. Como se pode constatar, houve grandes alterações ao plano de tarefas idealizado para o último semestre que são, em primeiro lugar, justificadas pela falta de aprovação e resposta dos representantes das ferramentas em análise para o desenvolvimento de testes de performance e, depois, pela realização de outras tarefas que não estavam inicialmente planeadas, como é o caso do desenvolvimento da aplicação móvel e do conector de integração com o MedicineOne. Assim, desenvolvimento iniciou com a criação da estrutura de *backend* de suporte aos serviços dos diferentes componentes do sistema com vista a, gradualmente, progredir para o desenvolvimento dos processos de negócio base de suporte às funcionalidades definidas para a aplicação móvel.

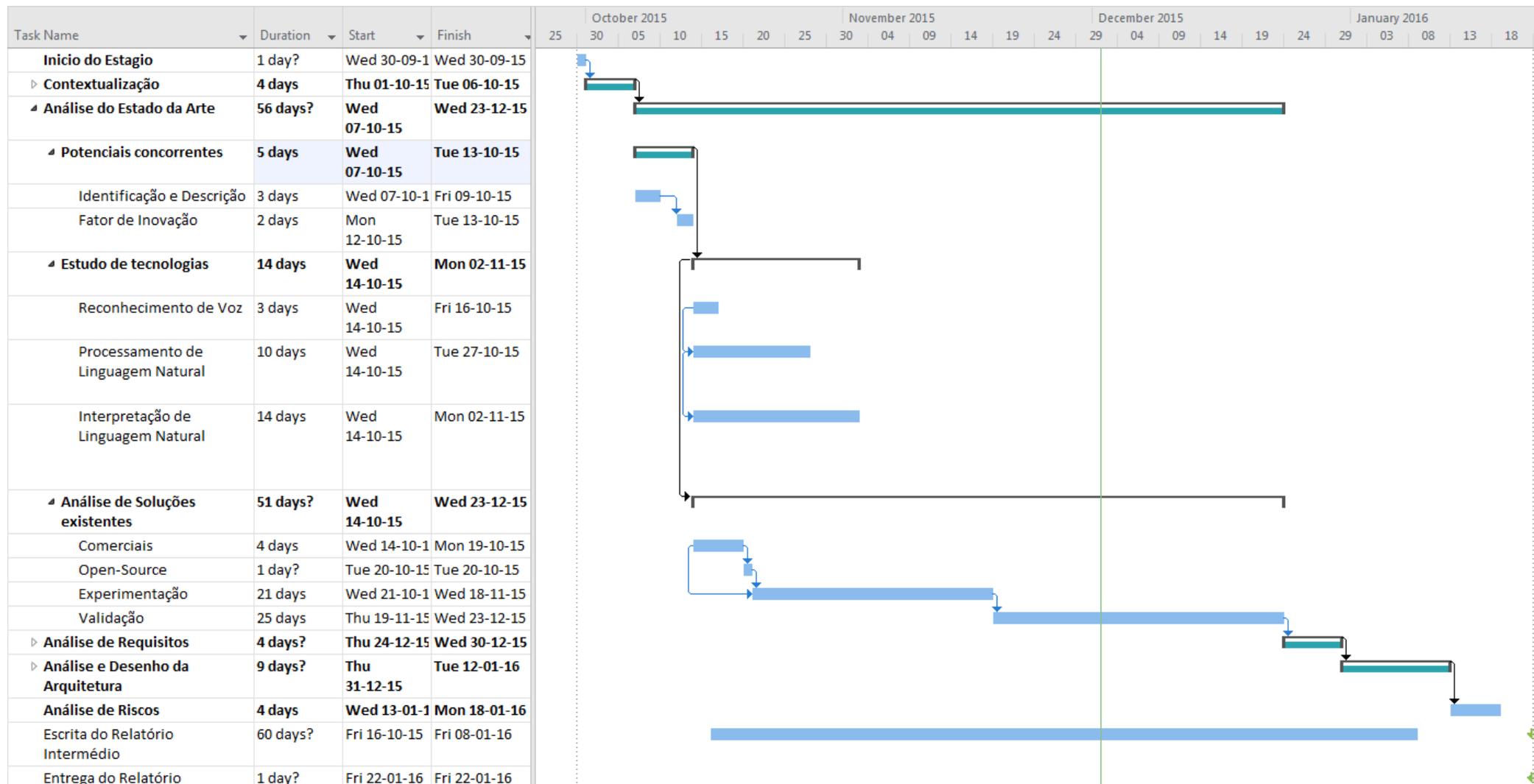


Figura 2 - Planeamento das tarefas referentes ao primeiro semestre

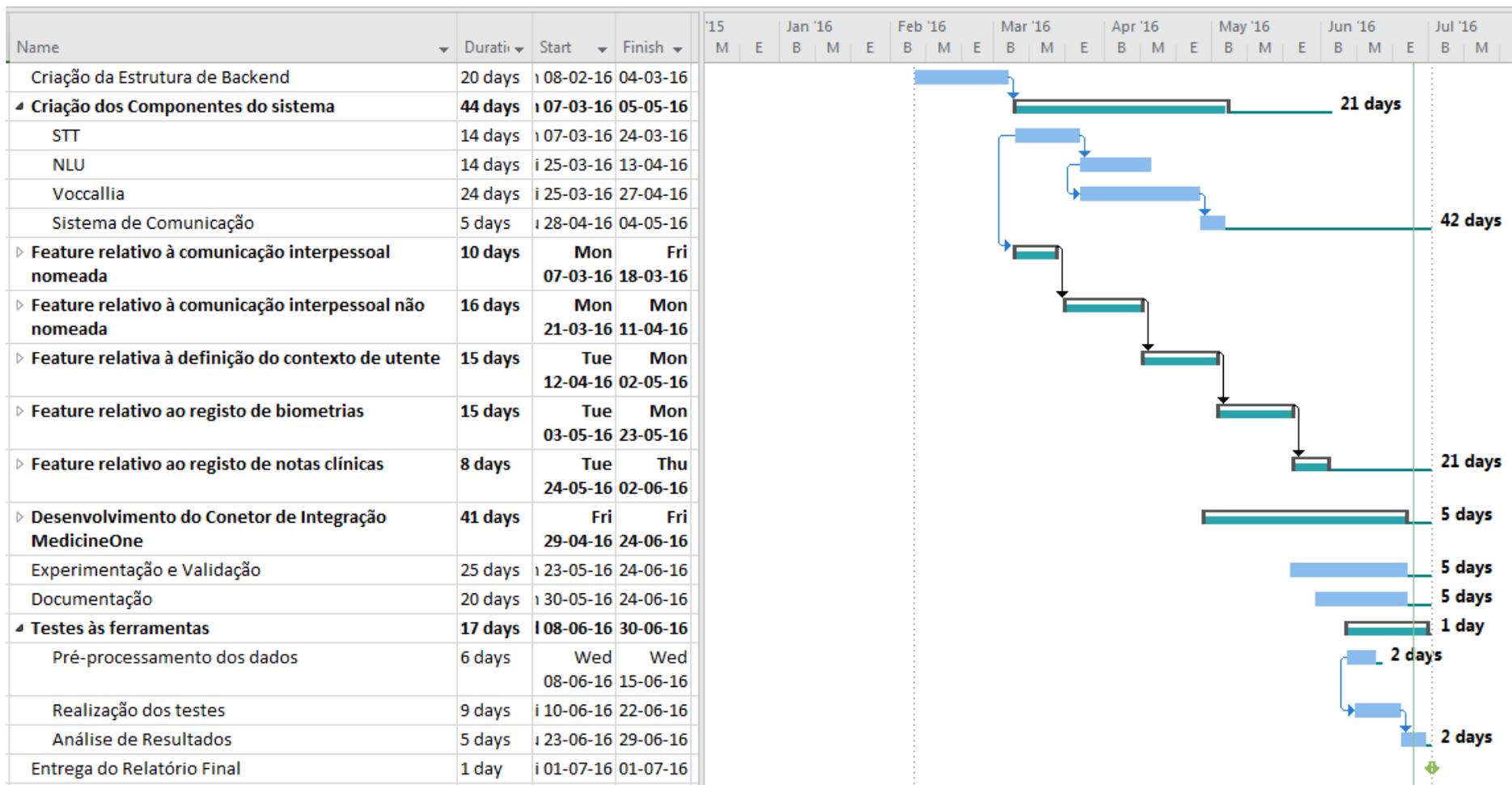


Figura 3 - Ilustração do plano de trabalhos seguido

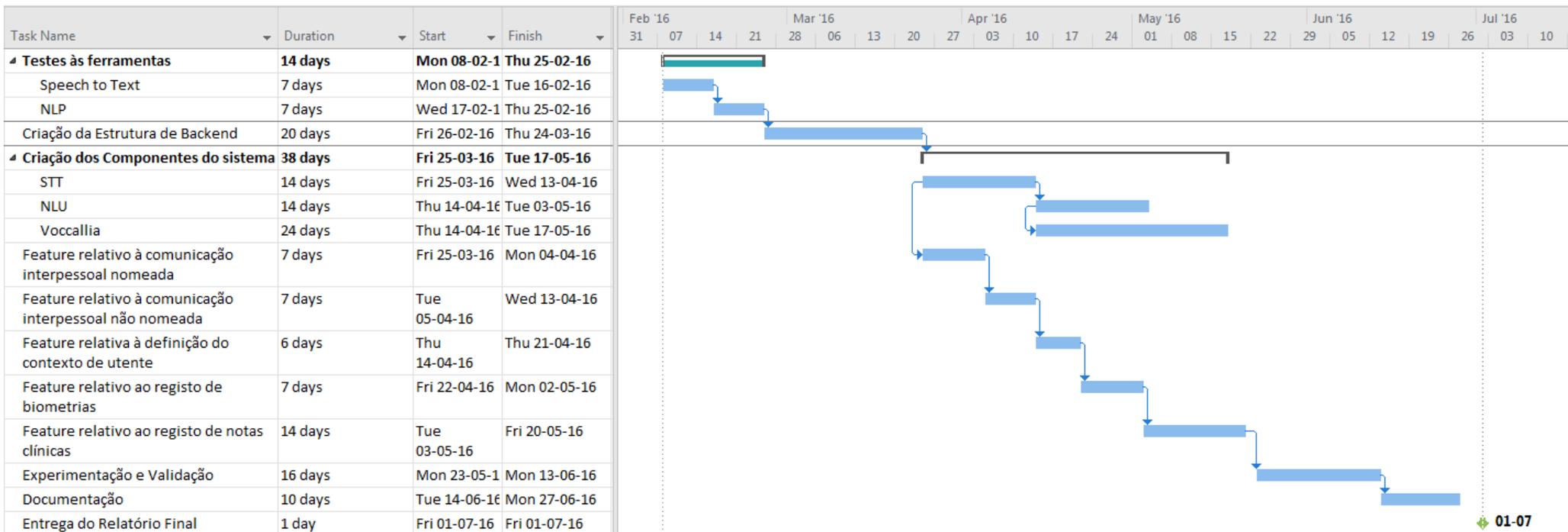


Figura 4 - Planeamento das tarefas referentes ao segundo semestre

2.2.4 Monitorização

Os processos e técnicas de monitorização no desenvolvimento de um projeto são um meio eficaz de gerir o esforço necessário para a realização das tarefas existentes. Além disso, permite também organizar, documentar e agilizar a tomada de decisões e dos processos de desenvolvimento relativos ao projeto.

Assim, as próximas subsecções têm o intuito de ilustrar os processos e técnicas de monitorização definidos para o âmbito do presente estágio.

Processos de Monitorização do Projeto

Neste subcapítulo são descritos os processos de monitorização utilizados para controlar o desenvolvimento e o cumprimento dos requisitos acordados. Alguns dos processos identificados servem de artefacto para o tipo de metodologia utilizada, sendo a sua utilização escolhida também de acordo com os processos já existentes no cerne da empresa. Desta forma, os processos de monitorização foram os seguintes:

- **Reuniões Internas** - com o intuito controlar possíveis alterações aos requisitos inicialmente acordados assim como novas formas de interação com o produto final que podem requerer a introdução de novos fluxos na arquitetura e execução dos serviços a serem desenvolvidos. Desta forma, devido à necessidade de se garantir estabilidade e consistência na especificação da arquitetura dos serviços, optou-se por estas reuniões se realizarem num período de duas em duas semanas. Assim, nas reuniões devem estar presentes as partes interessadas e necessárias para a fase específica do projeto (ex. **Workshops de Atributos de Qualidade** - Arquiteto, Estagiário e Cliente).
- **Reuniões com Orientador do DEI** - ainda que este não seja um estágio realizado para o Departamento de Engenharia de Informática e sim, no âmbito de uma empresa, foram realizadas reuniões com o orientador do DEI, o Professor Doutor Hugo Oliveira, uma vez que se achou que a sua experiência na área seria uma clara mais-valia na correta orientação do estagiário para a realidade e os desafios que caracterizam o âmbito de desenvolvimento do estágio.
- **Plataformas de comunicação interna** - de forma a tornar toda a comunicação entre as diferentes partes mais ágil, foram utilizadas as ferramentas internas. As ferramentas

utilizadas foram o *Skype for business* que se apresenta como sendo o meio mais fácil e eficiente para contactar as diferentes partes.

- **Microsoft Sharepoint** - esta plataforma é o principal meio de partilha e de organização da documentação relativa aos projetos desenvolvidos pela empresa. Por sua vez, e dado que esta plataforma permite a adição e personalização de *plugins* para novas funcionalidades, outros processos de monitorização são também realizados a partir daqui.
- **Controlo de versões e repositórios** - com vista de manter todo o trabalho realizado no estágio disponível e devidamente organizado internamente foi decidido utilizar o mecanismo de repositórios interno, **Team Foundation Server (TFS)** juntamente com o sistema **git**³. Assim, à medida que o trabalho é desenvolvido, pode documentar-se o estado de desenvolvimento de todas as *features* bem como eventuais problemas e **bugs** que possam vir a surgir durante o desenvolvimento, podendo depois ser resolvidos de acordo com o seu impacto e prioridade.

Técnicas de Monitorização

Dadas as características deste projeto, referidas na secção Objetivos do 1, optou-se por se definir diferentes técnicas de monitorização para cada uma das fases que contemplam o projeto. Na realidade, pode-se depreender do título de cada uma das secções que o trabalho relacionado com o **primeiro semestre** é muito mais direccionado para uma **vertente de investigação** enquanto no **segundo semestre** é um trabalho com uma **vertente mais de engenharia** - "processos estáticos/definidos".

Efetivamente, para o 1º semestre definiu-se o **diagrama de Gantt** como técnica de monitorização pela sua flexibilidade na capacidade de representação das tarefas e formas como se relacionam (**relações de precedência/dependência**) e na **interpretação rápida** da informação nele contido, ideal para avaliar o progresso das tarefas que compõem o plano de trabalhos. Note-se que, para o período relativo ao 1º semestre, estão previstos estudos de ferramentas para as etapas que caracterizam o projeto (secções que contemplam o capítulo seguinte - 3) com o objetivo de avaliar as suas potencialidades e eventuais limitações. Daí que, para a sua gestão, o diagrama de *Gantt* se torne naturalmente uma ferramenta ideal, dado que facilmente se pode observar se o tempo dedicado está ou não dentro do previsto.

³ <https://git-scm.com/>

Por outro lado, para o **2º semestre** para o qual está previsto o desenvolvimento dos serviços e criação do motor semântico, pretende-se que seja realizada uma monitorização mais orientada para o "ganho" de funcionalidades que se podem utilizar em conjunto com o protótipo da aplicação. Ou seja, dado que esta fase exige uma maior eficiência do esforço aplicado durante o desenvolvimento a técnica mais apropriada é a de **Gestão do Valor Ganho (EVM)**⁴. Com efeito, a principal vantagem desta técnica assenta no facto de que a partir da visualização do diagrama (custo atual/esforço realizado/tempo planeado) é fácil de perceber qual a situação do projeto no que diz respeito a "funcionalidades" implementadas e, assim, rapidamente adequar o esforço necessária para a conclusão de todas as tarefas existentes.

⁴ <https://www.mycpm.org/what-is-evm/>

3 Estado da Arte

No presente capítulo é realizada, em primeiro lugar, uma breve descrição do trabalho e das soluções já existentes no contexto do problema do presente projeto. Ou seja, sendo este projeto focado no modo de interação entre dispositivos móveis e pessoas, interessa compreender de que forma e quais as aplicações que conseguem promover este tipo de interação, onde através de comandos por voz são realizadas operações nos dispositivos e nos sistemas de informação que os suportam. Assim, e apesar de o foco de análise ser centrado em soluções que se inserem na área da saúde, não deixarão de ser apresentadas soluções que, embora fora do contexto atual, promovam características relevantes para o estudo e a realização do projeto atual.

Em segundo lugar, interessa saber quais as ferramentas existentes que permitem a realização das tarefas individuais necessárias para a criação deste tipo de sistemas. Para tal, irá então fazer-se a análise das soluções em dois mercados distintos: o mercado de soluções de reconhecimento de voz (ASR) e o mercado de soluções de compreensão de linguagem natural (NLU).

Por fim, a título de curiosidade, são descritas algumas das abordagens existentes na literatura para fazer face às motivações e objetivos identificados anteriormente (no capítulo 1) e, ainda, para demonstrar ao leitor a complexidade envolvida na conceção deste tipo de sistemas.

3.1 Abordagens na Literatura

A presente secção tem como objetivo descrever, de forma breve, as principais abordagens, na literatura relacionadas com a criação de um sistema com as capacidades de reconhecimento de voz e de compreensão de linguagem natural.

Com efeito, um sistema com estas capacidades pode ser descrito de acordo com Figura 5 em que são identificados e representados os seus componentes base. Estes compreendem as tarefas do reconhecimento de voz, compreensão do significado, até ao artefacto final que é a resposta por voz pelo sistema. Para o presente estágio, considerou-se que o desenvolvimento relacionado aos componentes de síntese de voz não é relevante para os objetivos do projeto.

Para a criação de um sistema de processamento e interpretação de linguagem natural, é necessária a especificação de mecanismos e formas de representação do conhecimento para diferentes as etapas. Este tipo de conhecimento tem sido muito estudado nos últimos 50 anos, existindo já um conjunto de modelos formais e ou teorias capazes de descrever a sua elaboração.

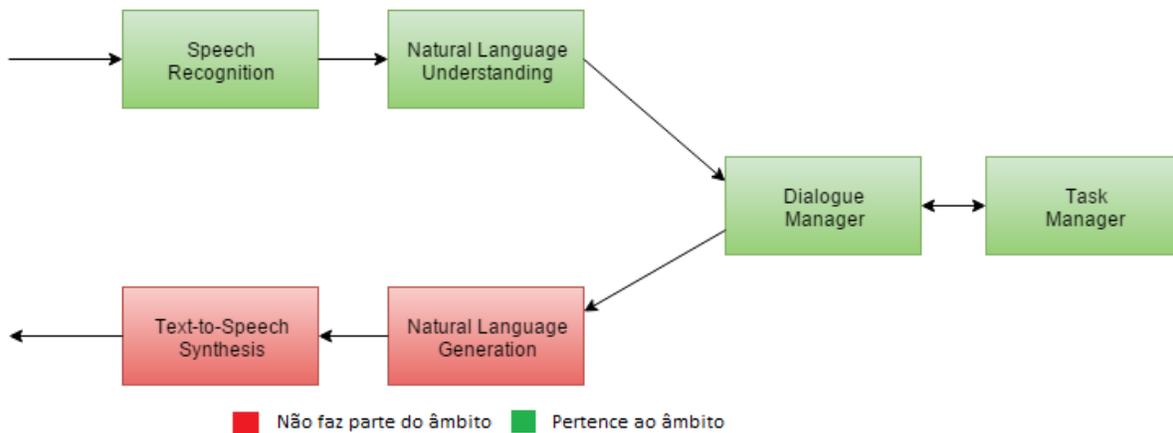


Figura 5 - Composição de um sistema conversacional, imagem adaptada de (Jurafsky & Martin, 2008)

Com efeito, todos estes modelos e/ou teorias podem ser representados recorrendo a **máquinas de estados** (autómatos), **sistemas de regras**, **lógica**, **modelos probabilísticos** e **de espaços vetoriais**. Da mesma forma, diferentes algoritmos podem ser utilizados para o mesmo efeito, destacando-se algoritmos de **Programação Dinâmica** e de **Classificação** no âmbito de **Machine Learning** (ML). De seguida, apresenta-se uma descrição superficial do funcionamento de cada uma destas técnicas e em que área de conhecimento são mais utilizados (Jurafsky & Martin, 2008).

Em primeiro lugar, máquinas de estados consistem no conjunto de símbolos/mecanismos entrada, dos estados e das transições entre si. As representações base são autómatos determinísticos e não determinísticos e transdutores de estados finitos. Associado a este tipo de representação, podemos ter sistemas de regras nas variações de **gramáticas regulares**, **gramáticas livres de contexto**, entre outras. Este tipo de representações são as principais ferramentas para a análise da **fonética**, **morfologia** e **sintaxe**.

Por outro lado, o uso da **lógica** está muito associada à modelação da semântica e da representação das relações de significado entre o objetivo do utilizador e a intenção - **pragmática** - ainda que, muito do trabalho desenvolvido mais recentemente, utilize técnicas mais robustas para extrair conhecimento a partir de semântica não lexical.

Em terceiro lugar, existe a possibilidade de se representar e capturar eficazmente as regras do domínio linguístico através da criação de **modelos probabilísticos**. Aliás, uma das grandes vantagens desses modelos é a possibilidade de se melhorar os modelos anteriormente referidos com a associação de probabilidades à captura dos valores de cada domínio. Na verdade, são vários os exemplos de outros modelos que podem ser melhorados com recurso às probabilidades, como é o

caso das máquinas de estados, que ao se associar probabilidades a cada um dos seus eventos (transições) tornam-se em **Modelos de Markov**. Estes modelos são especialmente importantes em tarefas nas quais existem problemas de **ambiguidade**.

Por fim, para lidar com os desafios que envolvem a desambiguação do sentido das palavras (Navigli, 2009) e o reconhecimento de entidade existe um conjunto mais específico de técnicas cujo objetivo consiste na análise global em termos semânticos das palavras e frases em vez da sua estrutura sintática. Para o primeiro caso, podemos referir algoritmos de computação de semelhança entre palavras (*word similarity*) ou a inclusão das relações com outras palavras, com por exemplo, **hiperonímia**, **hiponímia** ou a **meronímia**, presentes na **WordNet** (Fellbaum, 1998). A WordNet consiste numa base de dados lexical de Inglês onde nomes, verbos, adjetivos e advérbios são agrupados em estruturas de conhecimento, *synsets*, cada um representativo de conceitos distintos (About WordNet, 2016). Contudo, a WordNet não se limita à descrição/representação das palavras, mas sim, à relação entre os sentidos das palavras através das relações de hiperonímia, meronímia, antonímia, entre outros. Consequentemente, a existência de palavras com um grau elevado de proximidade entre si significa a possibilidade de serem semanticamente desambiguadas.

Por outro lado, para o segundo desafio, podemos referir a técnica de *semantic role labelling* ou *shallow parsing* que usa atributos da análise sintática com vista a estabelecer papéis semânticos na análise das partes com relação a determinado predicado. Para este efeito, definem-se os papéis de **Agent**, **Theme** e **Instrument** (Jurafsky & Martin, 2008).

Concluindo, esta é uma área com **elevadas dificuldades e expetativas** no que diz respeito às possibilidades de desenvolvimento futuro sendo, por isso, que **o âmbito deste estágio não prevê** o desenvolvimento de um sistema de raiz nas suas diversas componentes **mas a procura por soluções capazes de, numa ou mais componentes, fornecerem ferramentas capazes de ajudar a construir um sistema para ações de nível semelhante.**

3.2 Trabalho Relacionado

Atualmente, muito trabalho tem sido realizado na procura por novas interfaces de comunicação com os dispositivos móveis com o intuito de melhorar a experiência do utilizador e agilizar a realização de determinados processos que, manualmente, exigiriam muito mais foco da parte do utilizador. Uma destas interfaces e, talvez a mais usual nos dias de hoje, é a voz.

Na verdade, com a massificação dos *smartphones* e *tablets*, o uso de tecnologias como reconhecimento de voz e o processamento de linguagem natural no desenvolvimento de aplicações para este tipo de dispositivos tem sido uma constante. Porquê? Porque, hoje em dia, a *Internet* representa o maior meio de difusão e acesso a informação em tempo real disponível à escala global. Como tal, qualquer aplicação deste tipo está, naturalmente, ligada à *internet* já que cada vez mais isso representa uma necessidade e não uma qualidade. Assim, e sendo a voz um mecanismo natural e rápido de comunicação, muitas aplicações fazem uso dela para permitir a interação com os dispositivos com o intuito de libertar o utilizador para a realização de outras tarefas manuais enquanto interage por voz.

De fato, múltiplas aplicações têm sido desenvolvidas com o intuito de prover ao utilizador novas ferramentas e atalhos para lidar com as diferentes situações do dia-a-dia. Com efeito, para diversas situações já é possível com um *smartphone* reduzir o impacto que as mesmas têm na nossa vida ao verificar, por exemplo, qual o estado do tempo antes de sair de casa ou saber qual o caminho a tomar para o trabalho com o intuito de evitar acidentes ou trânsito uma vez que estas aplicações encontram-se ligadas aos diversos sistemas de informação e, assim, conseguem prover respostas rapidamente. Além disso, com o uso de sistemas *cloud* toda a informação passou a estar ainda mais acessível e com um menor tempo de acesso, tornando assim possível alimentar milhões e milhões de dispositivos com informação em “tempo real”⁵.

Desta forma, com evoluir e a maturação da tecnologia de reconhecimento de voz, diferentes áreas foram aparecendo no domínio de desenvolvimento de aplicações móveis e da IoT (*Internet of Things*), sendo os assistentes pessoais, os sistemas de documentação (narrativa e transcrição) e de integração as principais áreas de uso das tecnologias de reconhecimento de voz e processamento de linguagem natural. De seguida, apresentam-se os principais representantes de cada uma destas categorias.

⁵ Leia-se tempo próximo do real

3.2.1 Assistentes Pessoais: Google Now, Siri, Cortana, Assistant.ai e Personal Medical Advisor

Um dos tópicos de mais de maior disputa entre os principais fabricantes de sistemas operativos, em inglês *Operating Systems* (OS), para os dispositivos móveis é o nível de *hardware* e a performance associada ao sistema. Todavia, aquele que realmente tem vindo a demarcar a escolha entre os sistemas nos últimos anos é ao nível das aplicações e funcionalidades que cada um disponibiliza com o intuito de melhorar a experiência do utilizador.

Efetivamente, a tendência atual destas empresas (*Apple, Microsoft e Google*) tem-se demarcado com a apresentação de novas capacidades e funcionalidades nas suas imagens de marca: os **assistentes pessoais**, mais concretamente, a **Siri**, a **Cortana** e o **Google Now**, ilustrados na Figura 6 e o **Assistant.ai**⁶, na Figura 7. Com efeito, a aplicação *Assistant*, criada em 2010, conta já com mais de **20 milhões de utilizadores** e é a aplicação melhor classificada no âmbito de assistentes pessoais, competindo com as aplicações das principais marcas. Note-se que o sucesso da empresa é comprovado pela existência de colossos no desenvolvimento de equipamentos móveis e *software* relacionado, como é caso da **Motorola Solutions**⁷, **Intel Capital**⁸, **Alpine**⁹ e **SAIC Motor**¹⁰.

Com efeito, estes assistentes pessoais, na literatura definidos como **sistemas conversacionais** ou **sistemas de pergunta-resposta**¹¹, são cada vez mais uma aposta por parte das empresas envolvidas

⁶ <https://assistant.ai>

⁷ <https://www.motorolasolutions.com/>

⁸ <http://www.intelcapital.com/>

⁹ <http://www.alpine.pt/>

¹⁰ <http://www.saicgroup.com/English/>

¹¹ Consultar secção 3.1 para uma melhor descrição sobre este tipo de sistemas

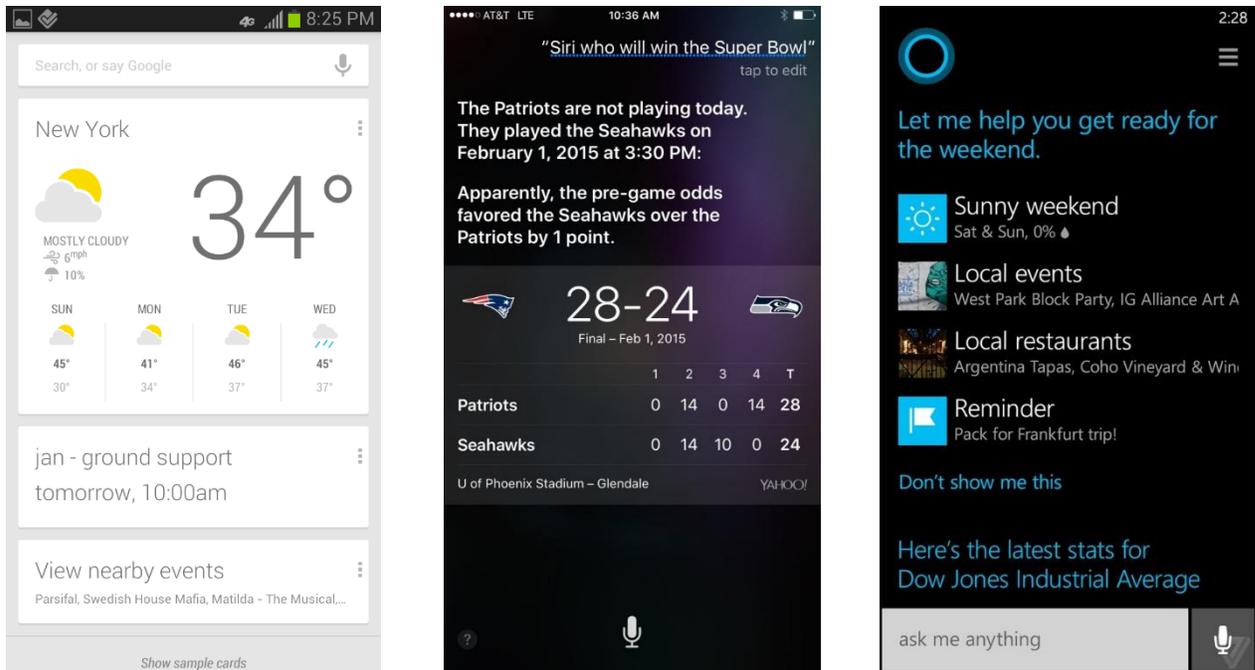


Figura 6 - Assistentes Pessoais dos principais sistemas operativos móveis - Google Now, Siri e Cortana

no desenvolvimento de sistemas inteligentes dado o aumento significativo de aplicações que este tipo de sistemas pode vir a suportar no futuro com o desenvolvimento de outras áreas de conhecimento. Em particular, pode-se referir o caso da domótica e dos sistemas de ajuda (ex. ajuda a idosos, ao ensino, telemedicina, entre outros), onde se podem destacar o mais recente produto da Google na área da domótica, o **Google Assistant**¹² (ilustrado na Figura 8), a tese de Mestrado *Personal Medical Advisor* (Lobo, 2013) realizado na Faculdade de Engenharia do Porto.

¹² <https://googleblog.blogspot.pt/2016/05/io-building-next-evolution-of-google.html>

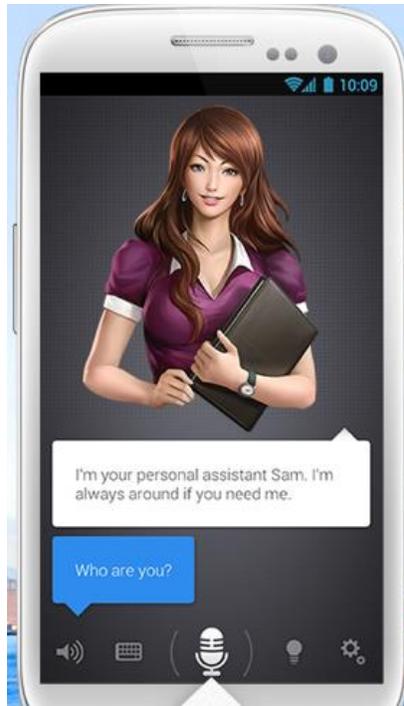


Figura 7 - Assistente Pessoal da api.ai - Assistant.ai

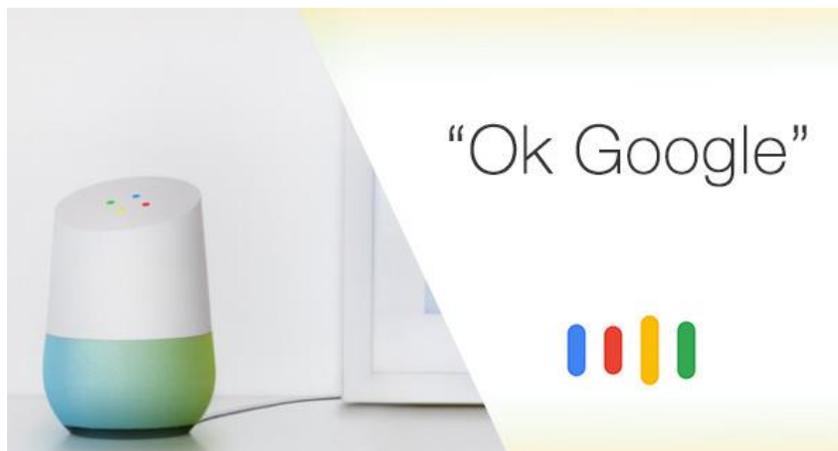


Figura 8 - Dispositivo Google Assistant apresentado no Google I/O 2016 no contexto da domótica

3.2.2 Sistemas de documentação: Nuance, Winscribe e VoiceInteraction

Hoje em dia, é muito comum e fácil encontrar-se em várias áreas de negócio, pessoas que necessitam frequentemente de introduzir ou registar dados no formato digital mas com uma maior eficiência. Ao contrário do que acontece com este tipo *sistemas*, o *software* que envolve o registo manual através do preenchimento de formulários com múltiplos campos é altamente ineficiente em termos de produtividade e, possivelmente, fonte de insatisfação da parte das pessoas que o usam.

Para além disso, este tipo de sistemas de documentação já se encontram muitas vezes integrados com dicionários de termos técnicos que permitem um aumento da eficácia no reconhecimento do discurso do utilizador, refletindo-se num aumento da produtividade do mesmo visto que não necessita de fazer edições do conteúdo com tanta frequência.

Assim, é possível ver uma listagem de exemplos de uso e de uma descrição breve de algumas ferramentas usadas em diversos contextos:

- **Sistemas de Ditado** – no contexto da saúde, existem várias ferramentas, sendo aquelas com maior destaque o *Dragon NaturallySpeaking*¹³ da Nuance e as ferramentas da Winscribe, o *Dictation, Mobile Suite e Text*.¹⁴ Estas ferramentas permitem a geração de documentos eletrónicos de contexto médico com mecanismos de enriquecimento do significado do conteúdo, como é o caso do suporte a termos e valores específicos de um domínio (ex. ICD-10¹⁵). Por outro lado, num contexto mais pessoal, podemos destacar a solução VoxControl da *VoiceInteraction*¹⁶, que é um sistema cujo propósito é de flexibilizar a interação com aplicações instaladas no sistema operativo *Windows* para a edição, formatação e, como não podia deixar de ser, ditado de documentos através da voz do utilizador.
- **Sistemas de Transcrição** – os sistemas de transcrição surgem associados a tarefas de automatização de trabalho que outrora teria de ser realizado manualmente. São exemplos de cenários de uso, a audição de notas clínicas por uma pessoa que, por sua vez, realizava a transcrição para um ficheiro de texto e, também, a legendagem automática de telejornais

¹³ <http://www.nuance.com/dragon/index.htm>

¹⁴ <http://idtfla.com/winscribe/>

¹⁵ <http://www.nuance.com/for-healthcare/resources/icd-10/index.htm>

¹⁶ http://www.voiceinteraction.pt/?page_id=1086

na imprensa. Neste contexto podemos destacar o Audimos.Server¹⁷ da *VoiceInteraction* e a plataforma *Clintegrity*¹⁸ da Nuance.

3.2.3 Sistema de integração: *Vocera*, Nuance

De fato, no universo de soluções disponíveis, tanto a empresa *Vocera Communications, Inc.*¹⁹ como a *Nuance Communications, Inc.* são referências na criação de soluções de comunicação na área da saúde. Por um lado, as forças motivadoras da *Vocera* centram-se na agilização e simplicidade de processos de comunicação entre de profissionais de unidades de saúde recorrendo a um conjunto de tecnologias bem consolidadas no *core* da empresa, como são as áreas de integração de *software* e de comunicação de dispositivos móveis. Desta forma, não é de espantar a vasta gama de serviços e soluções para os diferentes dispositivos (*tablets, smartphones, computadores*) e orientadas para a troca de informação entre profissionais de uma unidade saúde, sendo estas realizadas por voz, texto, alarmes e notificações. Assim, um dos aspetos vitais desta é empresa é a possibilidade de integração do sistema de informação de unidade de saúde com as suas soluções com o propósito de facilitar o acesso à informação em qualquer altura e lugar.

Por outro lado, no contexto da saúde, a *Nuance* tem uma abordagem mais virada para a disponibilização de novas funcionalidades centradas na obtenção e geração de conhecimento a partir de dados clínicos existentes, recorrendo a tecnologias de processamento de linguagem natural e de reconhecimento de voz. Com efeito, os avanços realizados pela *Nuance* em cada uma destas áreas, mais concretamente na melhoria do reconhecimento de voz, são bastante significativos. Com efeito, a capacidade de, hoje em dia, conseguir reconhecer a nomenclatura usada para a descrição de doenças, sintomas e outras expressões próprias do contexto da saúde é maioritariamente suportada pela empresa, tornando a informação resultante mais rica e útil para os utilizadores (ex. deteção automática termos clínicos aquando de uma transcrição ou de um ditado).

Na realidade, pode-se constatar a partir de uma análise do *website* de cada uma²⁰²¹, a vastidão de soluções criadas por cada uma das empresas. Contudo, restringindo a análise para o contexto do

¹⁷ http://www.voiceinteraction.pt/?page_id=374

¹⁸ <http://www.nuance.com/for-healthcare/clintegrity360/index.htm>

¹⁹ <https://www.vocera.com/>

²⁰ <http://www.nuance.com/products/index.htm>

²¹ <https://www.vocera.com/#>

presente projeto, as que mais se destacam pelas tecnologias e objetivo de utilização são as seguintes:

- ***Vocera Collaboration Suite*** – é uma solução de comunicação para dispositivos móveis (*smartphones*), integrada com os sistemas da unidade de saúde que fornece as capacidades de comunicação por voz, mensagem, notificação e distribuição de conteúdos com outras pessoas pelo reconhecimento por voz da identificação do nome, grupo de pertença de uma pessoa ou, simplesmente, por distribuição generalizada (*broadcast*) pelos utilizadores. É um produto em conformidade com os regulamentos *Healthcare Insurance Portability and Accountability Act* (HIPAA) cujo objetivo é o aumento da produtividade dos profissionais, assegurar a segurança de pacientes e garantia de uma experiência de cuidados. É, também, uma solução que tem o objetivo final de agilização dos fluxos de comunicação entre profissionais, de forma simples, eficiente e segura.
- ***Vocera Clinical Integrations*** – é uma solução de integração de sistemas de saúde com as funcionalidades e serviços disponíveis na *Vocera Platform* cujos casos de uso se estendem à integração com os programas de monitorização de utentes e das suas condições fisiológicas, gestão de camas, chamada de enfermeiros, segurança e comunicações de emergência, entre outros cenários. Assim como a solução anterior, esta tem como objetivo a agilização de processos dentro de uma unidade de saúde através da disponibilização de informação sensível no tempo para todos os dispositivos integrados, de forma segura e simples.
- ***Dragon Medical One*** – é uma solução de reconhecimento de voz direcionada para a agilização dos processos de documentação clínica diretamente relacionada com o aumento da produtividade pelos profissionais de saúde. Esta solução encontra-se em conformidade com os regulamentos de proteção de dados HIPAA e *European Data Protection Directive* (EDPD), como seria de esperar deste tipo de solução, onde são manipulados dados clínicos de pacientes e profissionais de saúde. Para além disso, é possível integrar a ferramenta com as soluções já utilizadas no contexto da unidade de saúde com vista a agilizar processos de documentação existentes.
- ***Nuance Healthcare Development Platform*** – é uma solução orientada para a integração com sistemas de informação na área da saúde. Compreende o uso das tecnologias de ASR e NLU com o intuito de possibilitar tanto o aumento da produtividade associada ao tratamento de paciente (através da documentação rápida e natural das condições que o

caracterizam) como o significativo aumento de qualidade dos dados que passam a ser integrados nos EHRs (é realizada a extração de informação a partir dos dados existentes – documentos, transcrições, etc. – com o intuito de fornecer uma maior representatividade e significado para o EHR, como é o caso da identificação de patologias e medicações em conformidade com os regulamentos e classificações próprios da área (ex. *SNOMED*, *RxNORM*).

3.2.4 Discussão

Como se pode constatar da análise realizada, existe um vasto leque de soluções nas diferentes áreas de negócio onde são utilizadas as tecnologias de reconhecimento de voz e de processamento de linguagem natural com o intuito de agilizar os processos (comunicação, acesso a informação ou execução de tarefas) e promover uma melhoria das condições de trabalho e satisfação dos profissionais nos seus locais de trabalho.

De fato, tomando o exemplo do *Vocera*, nomeadamente, a solução de integração clínica, *Vocera Clinical Integration*, é facilmente perceptível como múltiplos casos de uso pensados para a solução atual do Vocallia (ex. gestão de fluxo do utente, chamada de enfermeiros, segurança e chamadas de emergência) foram divulgados em casos de estudo como cenários de integração na conferência da HIMMS em Fevereiro do presente ano (Himms, 2016) como sendo suportados pela sua solução.

Para concluir, importa realçar a existência de mercado onde a solução atual pode de fato ser introduzida ainda que limitada pela concorrência pesada existente, como são o caso do *Vocera* e da própria *Nuance*, apesar de não incorporar soluções idênticas de integração pode no futuro entrar também no circuito com soluções mais viradas para o fluxo e agilização de processos. Contudo, tratando-se o mercado da língua portuguesa de um território ainda não conquistado nesta área, esta apresenta-se como uma oportunidade para a *MedicineOne* da mesma maneira que demonstram os casos de integração do *Vocera*.

3.3 Análise de ferramentas

Nesta secção é apresentada uma análise das ferramentas descobertas para a lidar com os desafios existentes do presente projeto. De notar que estes desafios estão associados, maioritariamente, às tarefas de reconhecimento de voz (ASR) e à compreensão da intenção do utilizador expressa em linguagem natural (NLU) sendo, desta forma, a análise centrada em ferramentas capazes de lidar eficazmente com estas questões, parcialmente ou na sua totalidade.

Para análise destas ferramentas definiram-se os seguintes critérios de comparação e hierarquização para as categorias de ferramentas em análise (ASR e NLU):

ASR (baseado nos critérios (Morbini, et al., 2013))

1. **Suporte à língua portuguesa no reconhecimento de voz.**
2. **Reputação associada à empresa e ferramenta.**
3. **Funcionalidades.**
4. **Modo de *deployment* da ferramenta** – se a ferramenta pode ser utilizada num domínio privado ou se tem de ser utilizado como serviço num domínio externo à empresa
 - a. **SaaS** – o serviço é *deployed* no domínio da empresa fornecedora do serviço e é disponibilizado para os consumidores através da *internet*.
 - b. **On-premise** – o serviço é *deployed* no domínio do consumidor do serviço.
5. **APIs de suporte a linguagens de programação, em inglês *Software Development Kit* (SDK)**

NLU

1. **Suporte à compreensão de entidades e frases na língua portuguesa** - expressas em linguagem natural.
2. **Modo de *deployment* da ferramenta** – se a ferramenta pode ser utilizada num domínio privado ou se tem de ser utilizado como serviço num domínio externo à empresa
 - a. **SaaS** – o serviço é *deployed* no domínio da empresa fornecedora do serviço e é disponibilizado para os consumidores através da *internet*.
 - b. **On-premise** – o serviço é *deployed* no domínio do consumidor do serviço.
3. **Reputação associada à empresa e ferramenta.**
4. **APIs de suporte a linguagens de programação (SDK)**

Com efeito, sendo o mercado de atuação da *MedicineOne* na sua totalidade em países pertencentes à Comunidade dos Países de Língua Portuguesa (CPLP)²² (com maior destaque para clientes em Portugal), o suporte à língua portuguesa²³ nas ferramentas de ASR e NLU é um requisito fundamental a ter em conta na altura de escolha de uma ferramenta.

Em segundo lugar, tratando-se a tarefa de ASR de uma tarefa vital para o sucesso de uma futura solução, importa assegurar que as ferramentas em análise tenham já uma posição marcada no mercado, para além, de suportarem o requisito de **suporte à língua portuguesa**. De notar que, o conceito de estatuto de uma empresa no mercado, por norma, tem relação direta com aumento dos custos associados ao uso da ferramenta. Como tal, este é um importante indicador uma vez que permite avaliar (positiva ou negativamente) o grau de confiança relativamente às potencialidades da ferramenta. Contudo, e como se pode constatar da hierarquia em cada uma das categorias, este critério surge em diferentes posições. Isto acontece, mais concretamente na categoria de NLU, devido ao fato de que, apesar de procurar sempre confiar no estatuto provido pela reputação como possível garantia de qualidade, estas podem não satisfazer as necessidades atuais do projeto pelas funcionalidades disponibilizadas. Além disso, muitas vezes ferramentas *open-source* ou de uso livre, apesar de sujeitas a determinadas condicionantes, como é o caso de licenciamento, conseguem fornecer qualidade de serviço e funcionalidade próximas das ferramentas “com maior reputação”.

De fato, outro ponto de avaliação de ambas as categorias de ferramentas inclui **as funcionalidades disponibilizadas** por elas. Isto é, no **contexto do reconhecimento de voz** relativo a uma frase ou comando dado pelo utilizador interessa, a quem está a querer usar a ferramenta²⁴, garantir que o que está a devolver ao utilizador como sendo o conteúdo da frase ou comando é aquele com um determinado grau de confiança. Ora, isto é, normalmente, conseguido de duas formas: em primeiro lugar, a ferramenta devolve o conteúdo reconhecido e o grau de confiança associado (calculado pela ferramenta); e, depois, com a devolução das possíveis alternativas associados ao reconhecimento da frase ou comando dado pelo utilizador, com os respetivos graus de confiança. Para ambos os casos, pode ficar a critério de quem esteja a usar a ferramenta fazer um processamento posterior às alternativas devolvidas com vista a devolver uma melhor sugestão ao utilizador. Ainda que, no último caso, este tenha um maior leque de opções para fazer o processamento. Além disso, quando

²² <http://www.cplp.org/>

²³ No contexto das tarefas de ASR e NLU importa destacar a especificidade do idioma a ser reconhecido ser o português europeu (PT-PT).

²⁴ Neste caso a empresa, *MedicineOne*, que quer disponibilizar a funcionalidade para os seus clientes.

se fala neste tipo de ferramentas, importa também estudar quais os formatos de áudio suportados. Isto porque, num contexto *mobile*, a escolha pelo formato de gravação pode influenciar a perda de informação no áudio resultante e, assim, diminuir a capacidade de reconhecimento do motor de voz. Tal fator pode ser, por vezes, eliminatório quando se tem em conta a gama do dispositivo que faz a gravação do áudio uma vez que este pode não possuir os requisitos necessários (*hardware* e *software*).

Por outro lado, **no contexto de NLU**, importa reconhecer quais as entidades e intenções do utilizador presentes na frase ou comando do utilizador de forma a conseguir, rápida e facilmente, parametrizar e reconhecer diferentes interações do utilizador com a aplicação e, ainda, mapear o resultado devolvido para uma estrutura de dados definida²⁵. Adicionalmente, são valorizadas ferramentas que incluam suporte às relações semânticas entre palavras e expressões (ex. relações de sinonímia, hiperonímia, inferência) uma vez que reforçam o suporte à língua pela diversidade de formas de expressão (próprias da língua portuguesa).

Para concluir, o **modo de deployment** é um ponto-chave no que diz respeito à avaliação de ferramentas com suporte ao *deploy on-premise* ou devidamente certificadas para o contexto de ação do presente projeto. Isto porque ele se enquadra no domínio da saúde, onde é frequente o uso e manipulação de informação sensível ao abrigo de determinados *standards* de proteção e de acesso a dados, como é o caso do *HIPAA*²⁶ e da *EDPD*²⁷, que variam de acordo com a localização onde o serviço se encontra em execução (Estados Unidos (U.S.) e União Europeia (E.U.)). Desta forma, a conformidade com estes *standards* é também um ponto essencial uma vez que, sendo muita da informação usada neste tipo de sistema sensível, torna-se necessário assegurar a sua proteção (ex. entidades associadas ao tratamento, pagamento e a operações realizadas). Como tal, a ferramenta que se venha a utilizar tem de, obrigatoriamente, suportar pelo menos uma destas necessidades.

Assim sendo, de seguida, é apresentada uma análise elaborada sobre as diferentes ferramentas.

Nuance

A Nuance, já apresentada na secção anterior, disponibiliza as seguintes ferramentas que se enquadram no contexto de desenvolvimento do presente projeto: *Nuance Mobile*, *Nuance Mix* e *o*

²⁵ Ver secção 3.1

²⁶ <http://www.onlinetech.com/resources/references/what-is-hipaa-compliance>

²⁷ *European Data Protection Directive* - <http://ec.europa.eu/justice/data-protection/>

360 Development Platform, ainda que apenas a primeira seja alvo de análise. Tal acontece pelo fato de o acesso às duas últimas ferramentas ser de acesso condicionado e sujeito a um processo de aprovação que, durante o período relativo ao estágio, não foi concluído.

O *Nuance Mobile* é um *Software Development Kit (SDK)* disponibilizado pela empresa para os desenvolvedores de aplicações móveis em torno do seu motor de reconhecimento de voz. O *SDK* é suportado para *Android*, *iOS* e, também, no formato de *endpoints HTTP*.

Relativamente às linguagens reconhecidas pelo motor de voz, o requisito fundamental é cumprido – suporte à língua portuguesa (PT-PT) – sendo, como tal, parte das ferramentas escolhidas para avaliação de performance no Capítulo 6, na secção de Testes Realizados. Para além disso, o facto de o serviço estar em conformidade com ambas as políticas de proteção de dados pessoais na área da saúde, mais concretamente a *EDPD*, faz com que a ferramenta seja legível de utilização em qualquer uma das suas alternativas de *deployment*.

Em termos de funcionalidades disponíveis, a Nuance desenvolveu o seu motor de voz para que, após o processamento do áudio enviado pelo utilizador, devolvesse o conjunto de N melhores resultados com maior grau de confiança. Assim, o programador fica com a possibilidade de escolha entre um posterior processamento a partir de estes resultados ou apenas devolver o resultado de maior confiança para o utilizador. Esta possibilidade dá-lhe uma maior liberdade durante o desenvolvimento da sua aplicação/sistema uma vez que não fica limitado a um resultado.

Para concluir, os formatos de áudio suportados são os seguintes:

- *Pulse Module Code (PCM)*
- *Speex*
- *Adaptative Multi-rate (AMR)*
- *Qualcomm Code-Excited linear prediction (QCELP)*
- *Enhanced Variable Rate Codec (EVRC)*

AT&T

A AT&T, mundialmente reconhecida pela sua supremacia na oferta de soluções no mercado das telecomunicações, tem também uma ferramenta disponível para a tarefa de ASR. A ferramenta encontra-se disponível no formato de *endpoint HTTP* com uma API de acesso e com SDK para as linguagens de Java, PHP, Ruby e C#.

Contudo, apesar da empresa ser devidamente reconhecida, as especificações da ferramenta disponibilizada são inferiores às restantes ferramentas analisadas. De fato, apesar de suportar os formatos essenciais para o processamento de áudio do utilizador (PCM, AMR, OGG e RAW) e ainda suportar as funcionalidades de devolução de hipóteses para o reconhecimento de voz com graus de confiança associados, as linguagens suportadas nativamente são limitadas. Na verdade, de forma a suportar outras linguagens, de entre as quais o português PT-PT, é necessária a especificação de uma gramática com a especificação de frases e palavras possíveis de reconhecimento.

Por outro lado, as questões relativas à disponibilização do serviço são semelhantes às dos restantes serviços na medida em que apenas se encontra disponível o formato SaaS que por sua vez não está em conformidade com os regulamentos de proteção de dados clínicos (EDPD / HIPAA).

Para concluir, apesar da solução estudada conseguir satisfazer o requisito fundamental de suporte ao idioma não representa uma solução viável para a empresa. Isto porque requer a criação e manutenção dos mecanismos de suporte à linguagem (especificação da gramática) além do fato de não estar disponível no formato *on-premise* que possibilitaria prover conformidade com os regulamentos de proteção de dados.

Google

A *Google* notoriamente reconhecida pela os avanços realizados nas diversas áreas de conhecimento e pela liderança em cada um dessas áreas, anunciou no início deste ano a intenção de disponibilizar para o “público” uma API de acesso ao seu motor de reconhecimento de voz com o intuito de lutar frente a frente com a *Nuance*. Com efeito, a capacidade de reconhecimento de voz da sua ferramenta é elevada sendo frequentemente usada em muitos dos seus produtos (secção 3.2.1) e pelos fabricantes de dispositivos móveis com o OS *Android*. Efetivamente, há um número muito reduzido de fabricantes que disponibilizam um sistema próprio (ex. S-Voice da Samsung). Assim, apesar de o acesso a esta API de demonstração²⁸ se ter tornado público recentemente pela empresa para os desenvolvedores de aplicações através da sua plataforma *cloud*, a que foi usada no contexto do projeto foi a API HTTP anterior. Esta opção deveu-se ao fato desta disponibilização ter ocorrido na fase final do estágio, impossibilitando uma análise mais profunda da mesma.

²⁸ <http://techcrunch.com/2016/03/23/google-opens-access-to-its-speech-recognition-api-going-head-to-head-with-nuance/>

Em segundo lugar, o requisito fundamental de suporte à língua portuguesa é cumprido. Como tal, é parte das ferramentas escolhidas para avaliação de performance no Capítulo 6, na secção de Testes Realizados.

Contudo, do que foi pesquisado o serviço não parece estar em conformidade com ambas as políticas de proteção de dados pessoais na área da saúde, mais concretamente a EDPD. Ainda assim, é possível a assinatura de um BAA (*Business Associate Agreement*) (Wolber, 2016) com o intuito de assegurar a sua conformidade e tornando, hipoteticamente, a ferramenta legível de utilização no formato de SaaS.

Em termos de funcionalidades disponíveis, a Google disponibilizou igualmente as funcionalidades em torno do seu motor de voz para que, após o processamento do áudio enviado pelo utilizador, seja devolvido o conjunto de N melhores resultados com maior grau de confiança.

Em termos de formatos de áudio suportados, a ferramenta usada apenas suporta apenas o formato de áudio *Flac*, de acordo com o especificado na tabela *Tabela 4*. Contudo, foi analisada a documentação da ferramenta *cloud* de reconhecimento de voz e esta já suporta uma gama mais vasta de formatos, apesar de a especificação dos mesmos ainda ser desconhecida. De entre os formatos suportados conhecidos estão os seguintes:

- *Pulse Module Code mu-law (PCMU)*
- *Linear-16 (LPCM)*
- *Adaptative Multi-rate (AMR)*
- *FLAC*

Microsoft Speech API & luis.ai

A plataforma luis.ai²⁹, direcionada para a vertente de NLU, encontra-se em **fase beta**. Juntamente com a Speech API, na vertente de ASR, fazem parte de um projeto que surge do trabalho realizado da equipa de investigação da **Microsoft** e do seu motor de pesquisa *web* – Bing. Atualmente, o projeto faz parte de um repositório maior de projetos – projectoxford.ai³⁰ - sob a gestão da **Microsoft** no qual estão compreendidas diferentes áreas de desenvolvimento e de interesse.

²⁹ <https://www.luis.ai/Home/About>

³⁰ <https://www.projectoxford.ai>

O objetivo desta iniciativa é o de permitir às aplicações mecanismos de compreensão de linguagem natural de forma rápida e eficiente. Na verdade, esta ferramenta tem como base o uso de modelos frequentemente utilizados e gerados para aplicações conhecidas, como é o caso do próprio Bing e da Cortana. Por outro lado, também é possível a criação de modelos especializados para diferentes áreas de conhecimento caso seja necessário. É de realçar que dada a natureza experimental já referida, não foi realizada uma análise aprofundada desta solução uma vez que não é assegurado o suporte ao reconhecimento de entidades e intenções em expressões na língua portuguesa.

Em termos de reconhecimento de voz, esta plataforma encontra-se limitada ao reconhecimento de um conjunto restrito de idiomas (inglês – US/GB, alemão, francês, espanhol, italiano e mandarim)³¹ sendo, desta forma, descartada no que diz respeito a testes de reconhecimento de voz. Por outro lado, em termos de funcionalidades disponíveis, o seu motor de voz também suporta a devolução do conjunto de N melhores resultados com maior grau de confiança usufruindo das potencialidades. Esta ferramenta apenas suporta um formato de áudio (PCM) e, dentro das ferramentas analisadas, é a única a limitar o áudio enviado a um conteúdo máximo de 10 segundos.

Relativamente ao funcionamento da ferramenta luis.ai, o modelo é bastante similar às plataformas já enunciadas. Através da especificação de entidades e de ações específicas é possível habilitar o reconhecimento semântico a uma aplicação. São utilizados mecanismos de *Machine Learning*, mais concretamente *Active Learning*³², para treino e reconhecimento de entidades e intenções.

Por fim, **comparativamente às ferramentas anteriormente apresentadas**, a ferramenta luis.ai apenas se encontra disponibilizada para comunicação HTTP e com suporte para um conjunto muito limitado de idiomas (inglês, espanhol, francês, italiano e chinês).

wit.ai

A wit.ai é uma plataforma *open source*, recentemente adquirida pelo **Facebook**³³³⁴ - 5 de janeiro de 2015 - cujo propósito é o de disponibilizar um conjunto de funcionalidades aos desenvolvedores de *software* que possibilitem a construção de interfaces de reconhecimento e interpretação de comandos de voz para aplicações e dispositivos móveis. Sendo o objetivo da plataforma a

³¹ <https://msdn.microsoft.com/en-us/library/mt613445.aspx>

³² http://hunch.net/~active_learning/

³³ <https://www.crunchbase.com/organization/wit-ai/#/entity>

³⁴ <https://wit.ai/jobs>

construção de aplicações "semelhantes" à *Siri*, as APIs desenvolvidas albergam as funcionalidades que uma solução direcionada para a vertente de NLU, ainda que com algum suporte para ASR.

Efetivamente, dado um *input* de voz para a plataforma (ou a sua transcrição) é realizada a descoberta da intenção (**ação que deseja realizar**) do utilizador e quais as entidades envolvidas. Isto é conseguido através de um algoritmo de ML no qual não são apenas detetadas palavras presentes nas instâncias, mas onde também é realizada a **semântica** sobre os conceitos relacionados com as mesmas (ex. reconhecimento temporal de **amanhã**, **próxima semana**) e o reconhecimento das entidades semânticas (ex. Temperatura → valor 70º, Local → o meu quarto -- cenário ilustrado na Figura 9).

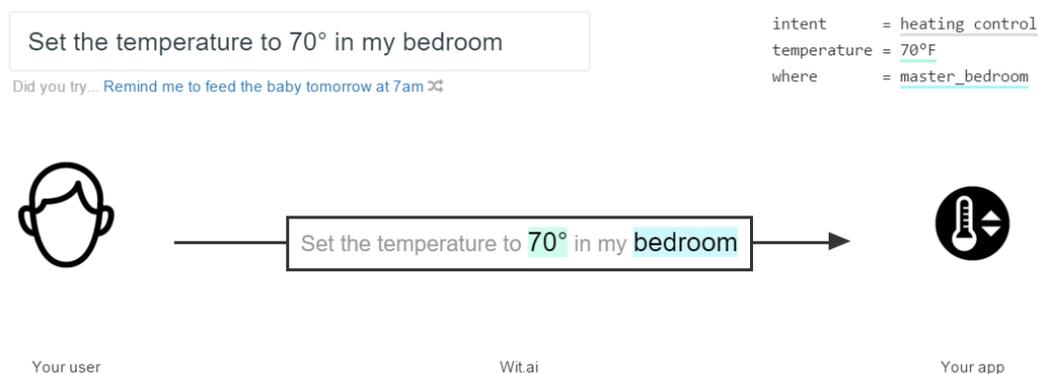


Figura 9 - Exemplo de reconhecimento de ações e entidades da plataforma wit.ai

Todo este reconhecimento é realizado a partir da especificação de um conjunto de instâncias de treino nas quais é realizado o processo de anotação das entidades e valores correspondentes pelo utilizador (**Aprendizagem supervisionada**). Assim, para se conseguir o melhor proveito dos mecanismos de classificação é necessária a especificação de um conjunto variado de instâncias/exemplos de treino para que o algoritmo consiga aprender para o contexto pretendido - (**corpus**). Adicionalmente, importa destacar que, para além das potencialidades referidas anteriormente, as funcionalidades expostas pela plataforma não albergam custos quer para os utilizadores quer para os desenvolvedores. Também, a possibilidade de interpretação múltipla de um *input* por parte do utilizador adjacente a um **nível de confiança** permite que diferentes interpretações estejam disponíveis em caso de incerteza do contexto no qual o utilizador se encontra inserido.

Em termos de *deployment*, tratando-se esta solução de um projeto *open-source*, seria apenas possível o uso da ferramenta tanto nas vertentes de *on-premise* já que o projeto não prevê qualquer tipo de regulamentação de proteção de dados, necessária para o projeto.

Desta forma, e com o intuito de abraçar uma maior comunidade, foram desenvolvidas APIs e SDKs compatíveis com as diferentes plataformas existentes, sendo as linguagens e tecnologias suportadas as seguintes:

- HTTP
- Python
- Node.js
- Ruby

Para concluir, é de realçar o fato de apesar de ser possível interagir com esta ferramenta com os formatos de áudio a seguir mencionados, o reconhecimento de voz por esta ferramenta encontra-se limitado a dois idiomas (inglês e francês), tornando como tal a ferramenta ilegível para a realização de testes nesta vertente. Os formatos suportados são:

- *Pulse Module Code (PCM)*
- *MP3*
- *Raw*
- *PCMU*

api.ai

Tal como a plataforma wit.ai, também esta, funciona através do uso de **algoritmos de ML** a partir da definição de um conjunto de instâncias que caracterizam as ações (**leia-se intenções**, para efeitos de concordância com a conceptualização da próprias plataformas) que o utilizador pode realizar (aprendizagem supervisionada). De facto, todo o funcionamento da plataforma é semelhante ao da plataforma anteriormente identificada, com a exceção de alguns apontamentos. Na generalidade, todas as funcionalidades providas pela wit.ai são também cobertas pela api.ai, diferenciando na forma como são definidas e não tanto na conceptualização (que é igual). Adicionalmente, a plataforma integra ainda outras formas de funcionamento que não são de momento suportadas na totalidade ou mesmo de todo pela wit.ai, como é o caso da **definição e transição de contextos** num ambiente conversacional, a criação de uma resposta pré-definida a uma ação reconhecida e o suporte a um vasto conjunto de domínios de conhecimento para resposta a ações do utilizador (ex.

resposta a questões de desporto, meteorologia, conhecimento geral, etc.). Por outro lado, são também suportadas as funcionalidades de semântica de relação de palavras de sinonímia com o intuito de diminuir a necessidade de especificação de diferentes formas de expressão de uma mesma intenção.

De facto, em termos de suporte a uma maior comunidade de utilizadores, as funcionalidades de ASR e de NLU encontram-se disponíveis para o conjunto de sistemas e linguagens na forma de SDKs e APIs presentes na Tabela 3.

Relativamente ao suporte ao nível de ASR, a plataforma fornece suporte ao reconhecimento da língua portuguesa (PT-PT) ainda que com capacidades limitadas uma vez que esta vertente de tecnologia não é o foco da empresa mas, sim, a vertente de NLU. Contudo, num plano de uso empresarial existe a possibilidade de aumento das capacidades de ASR com a criação de modelos de voz personalizados e adequados para o contexto de negócio no qual se pretende usar a ferramenta.

SDKs	APIs
<ul style="list-style-type: none"> • Assistant.ai • Facebook • Cortana • Slack • Kik 	<ul style="list-style-type: none"> • Twillio • Telegram • Alexa
	<ul style="list-style-type: none"> • Android • iOS • Ruby • Webkit HTML 5 • JavaScript • Node.js
	<ul style="list-style-type: none"> • Unity • .NET / C++ • Xamarin • Python • Cordova

Tabela 3 - APIs e SDKs suportados pela plataforma api.ai³⁵

Para concluir, o *deployment* dos serviços da ferramenta podem ser realizados tanto na vertente SaaS como *on-premise* possibilitando desta forma lidar com as questões de confidencialidade e proteção de dados que não estão previstas pela empresa e ao qual o contexto do presente projeto está inerentemente associado.

³⁵ <https://docs.api.ai/>

3.4 Discussão

Da análise das ferramentas efetuada na secção anterior é possível verificar que nem todas satisfazem os critérios de comparação definidos inicialmente. Mais concretamente, as ferramentas falham ora na capacidade de suportar a língua portuguesa nas tarefas de ASR ou de NLU, ora no requisito de garantia de conformidade com os regulamentos de proteção de dados (HIPAA / EDPD) necessários para o contexto do projeto e associados ao critério de *deployment* dos serviços.

Foi realizada uma avaliação comparativa entre as ferramentas, ilustrada na *Tabela 4* e na *Tabela 5* relativamente aos formatos de áudio suportados pelas ferramentas e aos critérios de suporte à língua portuguesa (PT-PT), modo de *deployment* e conformidade das mesmas com os regulamentos de proteção de dados (HIPAA e EDPD).

A avaliação relativa aos formatos de suporte de áudio permitiu-nos identificar quais as ferramentas que se encontram preparadas para dar suporte a uma gama mais variada de formatos de áudio. Isto porque, nos dias de hoje, diferentes dispositivos e sistemas operativos realizam gravações e transformações no áudio para formatos não *standard* (ex. *OGG, QCELP*).

	Taxa de Amostragem (kHz)	api.ai	AT&T	Google	IBM	Microsoft Speech API	Nuance
AMR		Não	Sim	Não	Não	Não	Sim
Flac	16000	Não	Sim	Sim	Sim	Não	Não
EVRC		Não	Não	Não	Não	Não	Sim
Ogg	8000	Não	Sim	Não	Não	Não	Sim
	11025	Não	Sim	Não	Não	Não	Sim
	16000	Não	Sim	Não	Não	Não	Sim
PCM	8000	Não	Sim	Não	Sim	Sim	Sim
	16000	Sim	Sim	Não	Sim	Sim	Sim
	22000	Não	Não	Não	Sim	Sim	Sim
QCELP		Não	Não	Não	Não	Não	Sim
Raw	8000	Sim	Sim	Não	Não	Não	Não
	16000	Sim	Sim	Não	Não	Não	Não

Tabela 4 - Formatos de áudio e taxas de amostragem suportadas por cada um dos motores de ASR

	Suporte PT-PT	Deployment		Conformidade com regulamento de confidencialidade	
		SaaS	On-premise	HIPAA	EDPD
Nuance	Sim	Sim	Sim	Sim	Sim
IBM Watson	Não	Sim	Não	Sim	Não
Google	Sim	Sim	Não	Não	Não
api.ai	Sim	Sim	Sim	Não	Não
wit.ai	Não	Sim	Sim	Não	Não
Microsoft Speech API	Não	Sim	Não	Não	Não
luis.ai	Não	Sim	Não	Não	Não
Audimos	Sim	N.D.	N.D.	N.D.	N.D
AT&T	Sim*	Sim	Não	Não	Não

* - Nestes casos é possível o suporte à língua através da definição de gramáticas de suporte (ex. pronúncia de palavras)

N.D. – Não Definido

Tabela 5 - Comparativo das ferramentas conforme os critérios definidos para o contexto da solução

Para concluir, verificou-se que as ferramentas legíveis de utilização e testes durante o período de estágio eram a da Nuance, Google e api.ai em termos de reconhecimento de voz e, a ferramenta api.ai e wit.ai para a vertente de NLU.

4 Análise e Especificação de Requisitos

I keep six honest serving men. They taught me all I knew. Their names are What and Why and When and How and Where and Who.

Rudyard Kipling

Este capítulo apresenta a abordagem usada para o levantamento dos requisitos funcionais e não funcionais a identificados no âmbito do estágio.

Com vista a especificar o conjunto de requisitos da solução, foi realizado em primeiro lugar uma análise dos potenciais *stakeholders*³⁶ com o intuito de analisar, devidamente, quais os interesses/vantagens que estariam na base de futuros cenários de uso da parte destes clientes.

Posteriormente, com a ajuda dos principais representantes internos e arquiteto de *software* da empresa, procurou-se validar as necessidades anteriormente identificadas com as necessidades de negócio e potenciais concorrentes nesta área com o intuito de identificar a sua viabilidade através potenciais oportunidades no mercado. Visto que no momento presente não existir ainda um cliente final, esta reunião interna permitiu definir um conjunto limitado de requisitos funcionais e não-funcionais base de forma a avaliar a robustez e viabilidade das tecnologias atuais para a construção de um sistema que permita a empresa entrar e singrar neste ramo.

³⁶ Principais clientes interessados na compra o uso da solução (Wikipedia, 2016)

4.1 Atores do Sistema

Nesta secção são identificados e descritos os principais atores que irão interagir com o sistema. É também identificada qual a responsabilidade que os mesmos têm na sua interação com o sistema.

Ator	Descrição	Papel
Paciente	Utilizador que interage com um dispositivo fixo.	Interage com o dispositivo com o intuito de efetuar pedidos/queixas para os quais necessita de intervenção por parte de um profissional de saúde.
Seguranças	Utilizador que interage com um dispositivo móvel (<i>smartphone</i> ou <i>weareble</i>).	Interage com o dispositivo com o intuito de realizar uma ação de resposta a intervenções ou de comunicação com profissionais.
Profissional de Saúde	Utilizador que interage com um dispositivo móvel (<i>smartphone</i> ou <i>wearable</i>).	Interage com o dispositivo móvel com o intuito de realizar uma ação de análise clínica, alteração das condições de físicas referentes a um utente ou de comunicação com outros profissionais.

Tabela 6- Identificação e descrição dos atores intervenientes no sistema

Sendo que o ator **Profissional de Saúde** pode ser decomposto nos seguintes atores:

Ator	Descrição	Papel
Médico	Utilizador que interage com um dispositivo móvel (<i>smartphone</i> ou <i>weareble</i>).	Interage com o dispositivo com o intuito de efetuar ações de análise clínica (ex. registo de notas clínicas, início do processo de alta de utente) e de comunicação com outros profissionais.
Enfermeiro	Utilizador que interage com um dispositivo móvel (<i>smartphone</i> ou <i>weareble</i>).	Interage com o dispositivo com o intuito de realizar uma ação clínica (ex. registo de biometrias, avaliação do processo de alta de utente) e de comunicação com outros profissionais.
Técnico Auxiliar de Saúde	Utilizador que interage com um dispositivo móvel (<i>smartphone</i> ou <i>wearable</i>).	Interage com o dispositivo móvel com o intuito de realizar uma ação alteração das condições físicas referentes a um utente (ex. prestação de cuidados de limpeza e alimentação ao utente) e de comunicação com outros profissionais.

Tabela 7 - Decomposição do ator Profissional de Saúde em outros atores

4.2 Requisitos Funcionais

A especificação dos requisitos funcionais para o presente projeto foi realizada através da utilização de *user stories* (Wiegers & Beatty, 2013). A sua utilização apresenta-se como sendo a mais indicada de acordo com a metodologia de desenvolvimento, *FDD*, definida para o presente projeto, uma vez que estas são descrições curtas e concisas (entre uma a três frases) nas quais se descreve, na perspetiva do utilizador final, as funcionalidades a implementar. De notar que, sendo esta descrição na perspetiva do utilizador, a linguagem utilizada é, normalmente, informal e simples com o intuito de tornar todo processo mais fácil de se perceber pelos utilizadores. Desta forma, a estrutura usual de uma *user story* é a seguinte (Wiegers & Beatty, 2013):

Enquanto <ator do sistema>

Quero <que seja cumprido um objetivo ou realizada uma ação>

Para que <determinado resultado seja refletido>

Com esta estrutura é facilmente perceptível que, para além do objetivo do utilizador que é comum à especificação em tabelas e diagramas de casos de uso, é possível entender qual a classe do utilizador e o racional da funcionalidade especificada. Por sua vez, isto permite que as *user stories* possam ser gradualmente refinadas e definidos os testes de aceitação das mesmas conforme as conversações vão progredindo junto do cliente. Por sua vez, a especificação em casos de uso que requer um maior detalhe na formalização dos requisitos e dos testes a realizar para cada caso de uso. Consequentemente, a escolha pela especificação em *user stories* acaba por ser a mais recomendada para a metodologias ágeis cujos projetos não necessitem de rigor nem formalização, dada a simplicidade e concisão que caracterizam a suas descrições (Wiegers & Beatty, 2013).

Por outro lado, apesar deste tipo de representação não fornecer uma especificação daquilo que são os requisitos para os serviços do sistema, permite entender quais os serviços exigidos pelo sistema com vista a fornecer a funcionalidade especificada à aplicação. É ainda possível validar qual o fluxo funcional que os serviços devem apresentar pela junção de diferentes *user stories* e pela validação da arquitetura desenvolvida.

Para finalizar a especificação dos requisitos, exige-se a realização da hierarquização dos requisitos de acordo com seu nível de prioridade para o desenvolvimento do projeto de forma a conseguir conciliar tanto o seu desenvolvimento como a posterior avaliação dos mesmos. Com este propósito, foi realizada a classificação dos requisitos de acordo com a nomenclatura descrita na Tabela 8.

Classificação	Significado
<i>Must</i>	O requisito especificado deve fazer parte das funcionalidades implementadas no final do projeto, sendo prioritária a sua realização.
<i>Nice</i>	O requisito especificado pode fazer parte das funcionalidades implementadas no final do projeto, sendo a sua implementação opcional.
<i>Whishful</i>	O requisito especificado foi identificado, mas não se prevê que exista capacidade para a sua implementação até ao final do projeto. O seu carácter é opcional.

Tabela 8 - Definição das classes de priorização dos requisitos funcionais

Com o intuito de permitir uma rápida perceção dos requisitos funcionais, a Tabela 9 apresenta um resumo dos requisitos especificados no âmbito do presente estágio.

User story ID	Descrição	Prioridade
B2C-01	Interação com a aplicação com vista a enviar uma mensagem a outro utilizador, nomeando-o por voz.	<i>Must</i>
B2C-02	Interação com a aplicação por voz por um profissional de saúde com o intuito de requerer a assistência de outros profissionais de saúde.	<i>Must</i>
B2C-03	Interação com a aplicação por voz por um utente com o intuito de requerer assistência de profissionais de saúde.	<i>Must</i>
B2C-04	Interação com a aplicação por voz por um profissional de saúde com o intuito de requerer o transporte de um utente a outro local.	<i>Nice</i>
B2C-05	Interação com a aplicação por voz por qualquer tipo de utilizador com o intuito de comunicar a ocorrência de distúrbios e/ou irregularidades.	<i>Nice</i>
B2C-06	Interação com a aplicação por voz por um Médico ou Enfermeiro com o intuito de definir um contexto de referência (cama de utente) para a realização da leitura de notas clínicas.	<i>Must</i>
B2C-07	Interação com a aplicação por voz por um Médico ou Enfermeiro com o intuito de definir um contexto de referência (cama de utente) para a realização da leitura de biometrias do utente.	<i>Must</i>
B2C-08	Interação com a aplicação por voz por um Médico ou Enfermeiro com o intuito de realizar o registo dos sinais vitais do utente.	<i>Must</i>
B2C-09	Interação com a aplicação por voz por um Médico ou Enfermeiro com o intuito de realizar o registo de notas clínicas adjacentes ao utente previamente definido.	<i>Must</i>
B2C-10	Interação com a aplicação por voz por um Médico ou Enfermeiro com o intuito de verificar a existência de camas vagas numa unidade de serviço.	<i>Nice</i>
B2C-11	Interação com a aplicação por voz por um Médico com o intuito de iniciar o processo de alta do utente.	<i>Whishful</i>
B2C-12	Interação com a aplicação por voz por um Enfermeiro com o intuito de dar continuação ao processo de alta do utente.	<i>Whishful</i>
B2C-13	Interação com a aplicação por voz por um Técnico Auxiliar de Saúde com o intuito de finalizar o processo de alta de um utente.	<i>Whishful</i>

Tabela 9 - Resumo dos requisitos funcionais levantados para o projeto

4.3 Requisitos Não-funcionais

A definição dos requisitos não-funcionais (ou atributos de qualidade) tem como principal objetivo a descrição das capacidades que o sistema deve compreender na sua concepção e funcionamento. Uma vez que este tipo de requisitos não expressa a funcionalidade esperada ou intenção do utilizador, seria errado a sua representação na forma de *user stories*. Assim, a sua representação foi realizada a partir de uma breve descrição do requisito, de cenários concretos da sua concretização e da descrição dos fatores, mensuráveis e visíveis, a analisar a partir do seu funcionamento. Mais concretamente, foi efetuada uma representação dos cenários concretos numa descrição em seis partes (Bass, Clements, & Kazman, 2003): a **fonte do estímulo**, o **estímulo**, o **Ambiente**, **Artefacto**, a **Resposta** e a **Medida da Resposta**. O seu significado é o seguinte:

- **Fonte do estímulo** – Entidade que gerou/realizou o estímulo.
- **Estímulo** – Condição necessária a considerar na chegada ao sistema.
- **Ambiente** – “Condições” nas quais o estímulo ocorre. O sistema pode apresentar diferentes condições aquando da ocorrência do estímulo.
- **Artefacto** – Parte do sistema que é estimulada. Pode incluir todo o sistema ou apenas algumas das suas partes.
- **Resposta** – Atividade realizada após a chegada do estímulo ao sistema.
- **Medida da Resposta** – Quando realizada, a resposta deve ser mensurável numa forma que o requisito possa ser testado.

4.3.1 Interoperabilidade

A presente solução deve ser concebida para que um futuro produto no universo da *MedicineOne* e que que pretenda fazer uso das funcionalidades de reconhecimento de voz e processamento semântico da informação possa facilmente comunicar com na presente solução a um custo reduzido, mantendo ainda a compatibilidade e suporte com os sistemas anteriores (sistemas *legacy*³⁷). De fato, a presente solução já deve prever um crescimento do número de aplicações (móveis ou *desktop*) e outros sistemas que deverão interagir consigo com o propósito de suportar a integração com diferentes sistemas de informação que não o da *MedicineOne* ou simplesmente de fornecer as funcionalidades já descritas.

³⁷ <https://www.techopedia.com/definition/635/legacy-system>

Como tal, o atributo de interoperabilidade é claramente um requisito primordial, uma vez que representa a capacidade de dois ou mais sistemas conseguirem trocar informações significativas entre si, de forma útil.

Título de cenário: integração com mecanismo de autenticação de um sistema externo ao ID:	
<i>Vocallia</i> RNI-01	
Atributo de Qualidade: Interoperabilidade	
Stakeholder que propôs a descrição do cenário de uso: <i>Product Owner, End User</i>	
Fonte(s) do Estímulo	Utilizador
Estímulo	Quer autenticar-se no <i>Vocallia</i> com um domínio pertencente a um EHR suportado.
Artefacto(s)	<i>Vocallia</i> e Conetor do EHR
Ambiente	A integração deve ser realizada em <i>Design Time</i> .
Resposta do Sistema	O utilizador é autenticado no <i>Vocallia</i> se as credenciais do utilizador forem validadas pelo EHR. Caso contrário, o utilizador não é autenticado no <i>Vocallia</i> .
Medida(s) de resposta	A informação relativa ao estado de autenticação e consequente acesso aos dados de sessão é devolvida ao utilizador num período de 3 a 5 segundos após realizado o pedido de autenticação.

Tabela 10 - Cenário concreto de execução do requisito de interoperabilidade 1

Título de cenário: integração com o repositório de pesquisa de profissionais de um EHR		ID: RNI-02
Atributo de Qualidade: Interoperabilidade		
Stakeholder que propôs a descrição do cenário de uso: <i>Product Owner, Developer</i>		
Fonte(s) do Estímulo	Utilizador	
Estímulo	Necessita de aceder a dados dos profissionais existentes no EHR e filtrá-los com base num conjunto de parâmetros.	
Artefacto(s)	Conetor do EHR	
Ambiente	A integração deve ser realizada em <i>Runtime</i> .	
Resposta do Sistema	O <i>Voccallia</i> recebe os resultados da pesquisa realizada pelo EHR e devolve ao utilizador.	
Medida(s) de resposta	A informação devolvida ao utilizador relativa à pesquisa deve ser de acordo com os parâmetros inseridos pelo utilizador e ser devolvida num período entre 2 a 4 segundos.	

Tabela 11 - Cenário concreto de execução do requisito de interoperabilidade 2

Título de cenário: integração com os processos de negócio externos ao sistema <i>Voccallia</i>		ID: RNI-03
Atributo de Qualidade: Interoperabilidade		
Stakeholder que propôs a descrição do cenário de uso: <i>Product Owner, End-user</i>		
Fonte(s) do Estímulo	Utilizador	
Estímulo	Inicia a uma ação de negócio no <i>Voccallia</i> . Este despoleta o processo de negócio correspondente e consequente integração com o EHR com base no fluxo de negócio interno ao <i>Voccallia</i> .	
Artefacto(s)	<i>Voccallia</i>	
Ambiente	A integração deve ser realizada em <i>Runtime</i> .	
Resposta do Sistema	O <i>Voccallia</i> devolve a resposta ao utilizador referente à entrada do processo de negócio no EHR. Se der entrada com sucesso no EHR, então o processo no <i>Voccallia</i> foi executado com sucesso.	
Medida(s) de resposta	A resposta à realização do processo de negócio deve ser fornecida ao utilizador num período de 2 a 4 segundos após o utilizador o inserir no <i>Voccallia</i> .	

Tabela 12 - Cenário concreto de execução do requisito de interoperabilidade 3

Título de cenário: integração com o repositório de pesquisa de utentes de um EHR		ID:
		RNI-04
Atributo de Qualidade: Interoperabilidade		
Stakeholder que propôs a descrição do cenário de uso: <i>Product Owner, Developer</i>		
Fonte(s) do Estímulo	Sistema externo ao <i>Voccallia</i>	
Estímulo	Necessita de aceder a dados dos utentes existentes de um sistema de informação e filtrá-los com base num conjunto de parâmetros.	
Artefacto(s)	<i>Voccallia</i>	
Ambiente	A integração deve ser realizada em <i>Runtime</i> .	
Resposta do Sistema	O <i>Voccallia</i> recebe os resultados da pesquisa realizada pelo EHR e devolve ao utilizador.	
Medida(s) de resposta	A informação devolvida ao utilizador relativa à pesquisa deve ser de acordo com os parâmetros inseridos pelo utilizador e ser devolvida num período entre 2 a 4 segundos.	

Tabela 13 - Cenário concreto de execução do requisito de interoperabilidade 4

Título de cenário: integração do <i>Voccallia</i> com o mecanismo de comunicação do EHR		ID:
		RNI-05
Atributo de Qualidade: Interoperabilidade		
Stakeholder que propôs a descrição do cenário de uso: <i>Product Owner, End-user</i>		
Fonte(s) do Estímulo	Utilizador	
Estímulo	Necessita de comunicar com entidades devidamente registadas e suportadas EHR com o intuito de lhes fazer chegar uma mensagem proveniente de um dos seus processos de negócio.	
Artefacto(s)	<i>Voccallia</i> , EHR	
Ambiente	A integração deve ser realizada em <i>Runtime</i> .	
Resposta do Sistema	O EHR devolve ao <i>Voccallia</i> informação relativa à entrega da mensagem nos destinatários da mensagem. O <i>Voccallia</i> devolve ao utilizador o resultado da operação.	
Medida(s) de resposta	A resposta à entrega da mensagem junto dos destinatários deve ser devolvida ao utilizador num período de 2 a 4 segundos.	

Tabela 14 - Cenário concreto de execução do requisito de interoperabilidade 5

4.3.2 Modificabilidade

Dada a natureza do projeto, é clara a necessidade do sistema ser capaz de acomodar alterações com o intuito de refinar ou corrigir eventuais erros e/ou falhas presentes em módulos, serviços ou processos. Para tal, é necessário que o custo dessa mudança não seja o principal fator impeditivo para a sua realização. Ou seja, o **requisito de modificabilidade** tem como base o custo de associado a determinada alteração e **refere-se à capacidade que um sistema de software tem para acomodar quaisquer alterações.**

Contudo, é evidente que a construção de um sistema perfeitamente modificável é algo impraticável em sistemas atuais devido à grande complexidade dos mesmos e, também, dada a necessidade em satisfazer outros requisitos de qualidade. Portanto, para se definir o nível de modificabilidade que o sistema deve suportar importa destacar quais as componentes do sistema que podem ser alvo de modificações. Para este efeito, analisou-se quais as necessidades futuras do sistema e de negócio da empresa e categorizou-se as componentes de acordo com o impacto que o requisito de modificabilidade teria em cada uma.

Deste modo, os diferentes níveis de modificabilidade encontram-se descritos através dos cenários concretos descritos nas tabelas a seguir apresentadas.

Título de cenário: Novo processo de negócio		ID: RNM-01
Atributo de Qualidade: Extensibilidade		
Stakeholder que propôs a descrição do cenário de uso: <i>Product Owner, End user</i>		
Fonte(s) do Estímulo	EHR	
Estímulo	Quer suportar processos de negócio customizados de acordo com o seu sistema informação e com o propósito de expor novas funcionalidades (ex. Processo de negócio de alta de paciente).	
Artefacto(s)	Vocallia (Orquestrador de processos)	
Ambiente	A alteração deve ser realizada em <i>Design Time</i> .	
Resposta do Sistema	O Vocallia recebe <i>input</i> para a realização do novo processo de negócio e devolve a resposta com a confirmação de execução e dados relativos para o utilizador.	
Medida(s) de resposta	O Vocallia reconhece o <i>input</i> como sendo direcionado ao processo de negócio adicionado e devolve a resposta de acordo.	

Tabela 15 - Cenário concreto de execução do requisito de extensibilidade

Título de cenário: Alteração do serviço de processamento semântico		ID:
		RNM-02
Atributo de Qualidade: Modificabilidade		
Stakeholder que propôs a descrição do cenário de uso: <i>Developer</i>		
Fonte(s) do Estímulo	<i>Product Owner</i>	
Estímulo	Há a necessidade substituir o serviço de processamento semântico que suporta o motor semântico do sistema <i>Voccallia</i> .	
Artefacto(s)	Ficheiro de configuração	
Ambiente	A alteração deve ser realizada em <i>Runtime</i>	
Resposta do Sistema	O componente de processamento de NLU deve efetuar realizar o processamento semântico com base no novo serviço.	
Medida(s) de resposta	O processamento semântico é realizado com base na nova implementação.	

Tabela 16 - Cenário concreto de execução do requisito de modificabilidade 1

Título de cenário: Alteração do serviço de reconhecimento de voz		ID:
		RNM-03
Atributo de Qualidade: Modificabilidade		
Stakeholder que propôs a descrição do cenário de uso: <i>Developer</i>		
Fonte(s) do Estímulo	<i>Product Owner</i>	
Estímulo	Há a necessidade substituir o serviço de reconhecimento de voz que suporta o sistema <i>Voccallia</i> .	
Artefacto(s)	Ficheiro de configuração	
Ambiente	A alteração deve ser realizada em <i>Runtime</i> .	
Resposta do Sistema	O componente de reconhecimento de voz deve efetuar a funcionalidade de ASR com base no novo serviço.	
Medida(s) de resposta	O componente substituído passa a realizar o reconhecimento de voz com base no novo serviço.	

Tabela 17 - Cenário concreto de execução do requisito de modificabilidade 2

4.3.3 Testabilidade

Garantir a robustez e validade das partes integrantes de um sistema é um **requisito crítico** quando se pretende avaliar a existência de falhas e de, que forma, estas podem voltar a ocorrer. Assim, se este não for devidamente pensado e especificado, o sistema construído dificilmente terá garantias de que a sua especificação não tem qualquer tipo de falhas durante o seu funcionamento.

Desta forma, e sendo este um sistema com múltiplos componentes e processos, é **imperativo** a existências de mecanismos de testes para garantia da qualidade do sistema.

5 Análise e Desenho da Arquitetura

The software architecture of a computing system is the set of structures needed to reason about the system, which comprise software elements, relations among them, and properties of both.

Anthony J. Lattanze

O presente capítulo descreve a arquitetura do sistema. Para este efeito, realizou-se a divisão do mesmo em quatro partes: **Identificação das Necessidades de Documentação, Descrição da Visão Global do Sistema, Decisões Arquiteturais e Arquitetura Desenvolvida.**

A arquitetura de um sistema é parte fulcral para o desenvolvimento de um projeto, uma vez que serve de modelo para o sistema e, ainda, para o próprio projeto que o desenvolve. Isto porque é na arquitetura que são definidas as atribuições de trabalho e é assegurado o cumprimento dos atributos de qualidade, sendo a sua estruturação e documentação peças chave para a sua manutenção pós-implantação. De realçar que este último passo será seguramente dos mais críticos para o sucesso de um projeto uma vez que o mesmo representa o meio comum de comunicação entre o arquiteto de hoje e o de amanhã.

Desta forma, a documentação deve ser realizada de acordo com o objetivo de permitir uma rápida e melhor perceção às pessoas envolvidas sobre como o sistema deve ser desenvolvido e dos cuidados a ter em conta, de acordo com suas as necessidades e responsabilidades dentro do projeto (Clements, et al., 2003). Esta análise encontra-se na secção **Identificação das Necessidades de Documentação.**

Posteriormente, na secção **“Descrição da Visão Global do Sistema”** é descrito qual o comportamento geral do sistema para fazer face aos requisitos definidos no capítulo **“Análise e Especificação de Requisitos”**. Aqui é dada uma perceção ao leitor sobre quais os componentes da arquitetura global do projeto *Vocallia*.

A secção de **“Decisões Arquiteturais”** tem como objetivo apresentar e conduzir o leitor sobre quais as principais decisões arquiteturais que se tomaram no decorrer do presente projeto com vista a fazer face aos requisitos definidos e a lidar com os desafios por eles apresentados. Desta forma, torna-se possível para o utilizador entender quais os princípios e pressupostos que estiveram na origem do problema e qual o impacto que as mesmas tiveram no desenho da arquitetura.

Por fim, na secção “**Arquitetura Desenvolvida**”, é apresentada a solução arquitetural desenvolvida para lidar com as funcionalidades e requisitos definidos para o período decorrente do estágio. Aí é demonstrada a arquitetura numa perspetiva mais detalhada uma vez que é necessário explicar todos os mecanismos utilizados na arquitetura para satisfazer as necessidades impostas para a solução.

5.1 Identificação das Necessidades de Documentação

Com efeito, a arquitetura deve prever diferentes níveis de detalhe para as diferentes pessoas que estão envolvidas num projeto de *software* (gestor de projeto, gestor de produto, programador mobile, programador *backend*, etc.). Isto porque diferentes necessidades de documentação surgem de acordo com as responsabilidades dentro do projeto ou de outros projetos com este (ex. integração de novos sistemas de informação com este sistema; desenvolvimento de aplicações móveis com base nas funcionalidades expostas, etc.) (Clements, et al., 2003).

Tendo em conta os objetivos anteriormente delineados para este sistema onde a necessidade de produção de documentação é um ponto vital para o sucesso do estágio, verificou-se a necessidade de estudar quais os tipos de documentação que deveriam ser produzidos na fase inicial do projeto e para quem ela deveria ser dirigida.

É evidente que, no desenvolvimento de um projeto de *software*, a criação e evolução da documentação deve sempre acompanhar os diferentes cargos e pessoas existentes no projeto e, como tal, mantida devidamente atualizada. Note-se que qualquer imperfeição na documentação pode implicar a falha de processos e de especificação, assim como o incumprimento na entrega da implementação das funcionalidades pedidas. A Tabela 18 é uma descrição dos principais *stakeholders* e das necessidades gerais de documentação definidas para o presente projeto. De notar que, apesar de representados todos os intervenientes e suas necessidades de documentação, o foco do presente estágio é apenas direcionado às necessidades “imediatas”³⁸ do projeto que, volto a realçar, são de suporte à integração, manutenção e extensibilidade do sistema *Vocallia*.

Por fim, na Tabela 19 é apresentada uma descrição das necessidades específicas dos *stakeholders* de foco para o presente estágio.

³⁸ O sentido **imediato** refere-se à orientação das necessidades de acordo com o impacto num futuro próximo no que diz respeito ao suporte do desenvolvimento e integração de funcionalidades junto com outros projetos.

	Vistas de Módulos					Vistas C&C	Vistas de Alocação				Outra Documentação					
	Decomposição	Uso	Generalização	Camadas	Modelo de Dados	Diversos	Deployment	Implementação	Instalação	Atribuição de Trabalho	Documentação de	Diagramas de Contexto	Mapeamento entre	Guias de Variabilidade	Análise de Resultados	Racional e Restrições
Gestor de Projeto	a	a		a			d			d		g				a
Membros da equipa de desenvolvimento	d	d	d	d	d	d	a	a	d		d	d	d	d		a
Analistas	d	d	a	d	d	a	d		a		d	d		a	d	a
Testers e Integradores	d	d	d	d	d	a	a	a	a		d	d	a	d		a
Product-Line Application Builders	d	d	a	g	a	a	a	a	a		a	d	a	d		a
Utilizadores finais						a	a		g							a
Novos Stakeholders	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
Equipa de Infraestrutura	a	a			a	a	d	d	g							
Clientes							g			g						a
Arquitetos de outros sistemas					a						d	g				
Arquitetos atuais e futuros	d	d	d	d	d	d	d	a	d	a	d	d	d	d	d	d
Manutenção	d	d	d	d	d	d	a	a			d	d	d	d		d
Chave: d = informação detalhada; g = informação generalizada; a = alguns detalhes; x = tudo																
■ Foco da documentação a ser produzida																

Tabela 18- Stakeholders e necessidades de documentação para o projeto Voccallia

	Vistas de Módulos					Vistas C&C				Vistas de Alocação			Outra Documentação				
	Decomposição	Uso	Generalização	Camadas	Modelo de Dados	Processos de Comunicação	Partilha de dados	SOA	Orquestração (Event-Based)	Deployment	Implementação	Instalação	Documentação de Interfaces	Diagramas de Contexto	Mapeamento entre vistas	Guias de Variabilidade	Racional e Restrições
Membros da equipa de desenvolvimento	d	d	d	d	d	d	d	d	a	a	a	d	d	d	d	d	a
Testers e Integradores	d	d	d	d	d	a	d	d	d	a	a	a	d	d	a	d	a
Arquitetos de outros sistemas						a	g	g					d	g			
Arquitetos atuais e futuros	d	d	d	d	d	d	d	d	d	d	a	d	d	d	d	d	d
Manutenção	d	d	d	d	d	g	d	d	d	a	a		d	d	d	d	d
Chave: d = informação detalhada; g = informação generalizada; a = alguns detalhes; x = tudo																	
 Documentação Obrigatória; Documentação necessária mas não obrigatória																	

Tabela 19 - Stakeholders e necessidades de documentação específicas para o âmbito do estágio

5.2 Descrição da Visão Global do Sistema

A presente secção pretende introduzir ao leitor a arquitetura idealizada para a resolução do problema do presente estágio através de uma primeira análise às principais entidades envolvidas no funcionamento do sistema e à forma como as mesmas interagem entre si. Assim, a análise irá realizar-se de acordo com a seguinte estrutura:

Em primeiro lugar importa identificar quais os principais agentes e de que forma interagem entre si. Assim, a partir da Figura 10 podemos visualizar o funcionamento global do sistema como este deve interagir com as aplicações que o rodeiam.

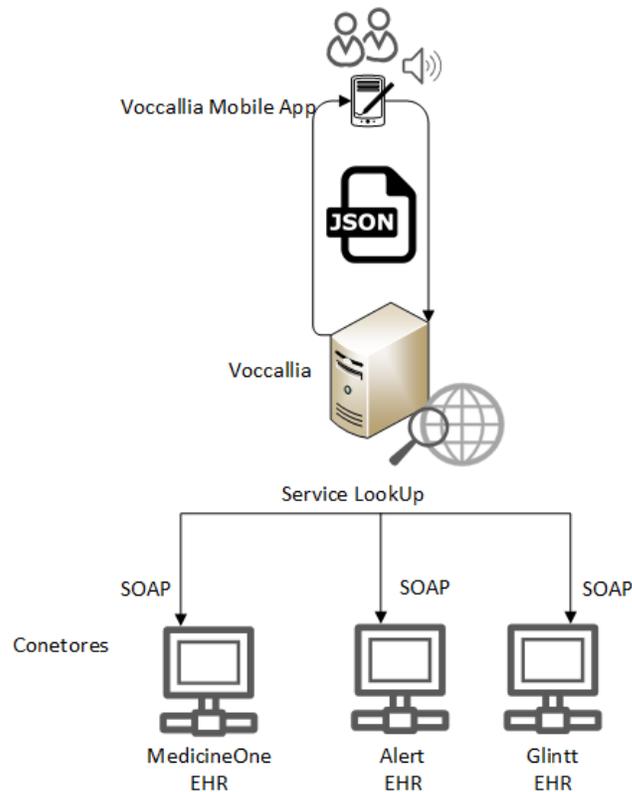


Figura 10 - Visão global da arquitetura do sistema Voccallia

Voccallia Mobile App é o *interface* responsável pela gestão e apresentação das funcionalidades disponíveis no sistema *Voccallia*. Para o cenário do presente estágio apenas se definiu a integração com uma aplicação **Android**, responsável pela realização dos requisitos funcionais especificados na secção 4.2. No momento de realização de uma ação, o áudio recolhido pela aplicação e, posteriormente, enviado para o *Voccallia*. A comunicação entre os dois é realizada via REST, sob o protocolo HTTPS.

De realçar que, no lugar da aplicação móvel pode ser inserido qualquer outro sistema uma vez que os mecanismos de interação entre ambos os sistemas são genéricos ao ponto de possibilitarem esse nível de integração.

Voccallia é o sistema de *backend* que se encarrega de fazer a descoberta da ação pretendida pelo utilizador a partir do áudio recebido e de lhe enviar a resposta para a ação reconhecida correspondente junto com toda a informação necessária para o utilizador efetuar a sua confirmação (ex. para a definição do contexto de análise de um paciente, deve ser enviado pelo sistema informação sobre os pacientes relacionados com o contexto do utilizador – cama em que se encontram, quarto, unidade de serviço, etc.). Por sua vez, após confirmada a ação por parte do

utilizador esta é comunicada ao sistema de informação/EHR ao qual ele se encontra atualmente associado (pelo *login*).

Conetor é o componente referente à implementação específica do mecanismo de integração de cada EHR com o *Voccallia*. De realçar que estes componentes são peças externas ao sistema. A comunicação é realizada através de SOAP/XML.

5.3 Decisões Arquiteturais

Como já referido anteriormente a presente secção tem como objetivo guiar o leitor pelas principais decisões arquiteturais tomadas na conceção e desenvolvimento da arquitetura do sistema do projeto *Voccallia*.

Efetivamente, aqui são apresentados os **princípios arquiteturais e estratégias** consideradas para melhor fazer face aos requisitos identificados e devidamente definidos no capítulo anterior. Note-se que, exatamente como o autor do excerto que inicia a presente a secção refere, o trabalho de um arquiteto não é a tomada de decisões referentes às tecnologias, bases de dados e ou ferramentas que agilizam o desenvolvimento e testes do sistema. Pelo contrário, **as decisões realmente importantes** são aquelas que tornam a **solução independente e devidamente isolada** das tecnologias que se poderão vir a usar aquando do desenvolvimento do sistema que, para o contexto da arquitetura, **são irrelevantes**.

Desta forma, a criação e desenvolvimento da arquitetura do sistema envolveu a tomada das seguintes decisões.

1. *Service-Oriented-Architecture (SOA)* (Hohpe & Woolf, 2004)
 - a. *Loose Coupling*
2. *Dependency Injection (DI)*
3. *S.O.L.I.D.* (Martin, 2002)
 - a. *Single Responsibility Principle (SRP)*
 - b. *Open/closed principle (OCP)*
 - c. *Liskov substitution principle (LSP)*
 - d. *Interface Segregation Principle (ISP)*
 - e. *Inversion of Control (IoC)*

5.4 Arquitetura desenvolvida

Nesta secção é realizada uma descrição mais detalhada da arquitetura desenvolvida para a presente solução e demonstrar como foram alcançados os requisitos não-funcionais definidos na secção 4.3, ilustrando quais as estratégias utilizadas, assim como as considerações tomadas de acordo a

respeitar a relação de prioridade entre requisitos. Para este efeito, são descritas as táticas arquiteturais e *design patterns* utilizados para cada um dos requisitos. Para a descrição da arquitetura utilizaram-se as técnicas e notações de representação presentes no livro (Clements, et al., 2003). Desta forma, para descrição da arquitetura utilizou-se as seguintes estratégias:

- **Descrição da organização do sistema** – Vista em Camadas
- **Descrição de comportamento em *Runtime* do sistema** – Vista Componente-e-Conetor
- **Descrição da análise do impacto de alterações no sistema** – Vista de Módulos, de Camadas e Componente-e-Conetor.

Para além disso, é apresentado o modo de funcionamento dos componentes chave do sistema *Vocallia*, como é o caso do mecanismo de orquestração de processos de negócio e a forma como o sistema interage com os diferentes EHRs.

Desta forma, esta secção encontra-se organizada de forma a descrever a arquitetura desenvolvida através da definição da sua estrutura e organização, identificação dos principais componentes existentes e, por fim, da descrição do comportamento esperado para as diferentes situações que caracterizam o sistema, mais concretamente, situações nas quais é possível comprovar como os atributos de qualidade foram alcançados, tendo em conta o seu nível de prioridade.

5.4.1 Estrutura e Organização

Sendo a presente solução baseada numa arquitetura orientada a serviços, como referido na secção anterior, importa conhecer de que forma o sistema se encontra organizado, identificando quais as camadas existentes e de que forma interagem entre si. Assim, na Figura 11 podemos ver a decomposição do sistema a partir de uma vista em camadas. Como consequência, é facilmente perceptível quais as principais camadas existentes, assim como a hierarquia de relações entre os seus constituintes e com as outras camadas. Por outro lado, esta vista permite também visualizar o isolamento de responsabilidade cada uma das unidades e suas camadas que é uma das características essenciais dos princípios definidos na secção 5.3 que, por sua vez, reflete positivamente o requisito de modificabilidade e extensibilidade necessários para este projeto.

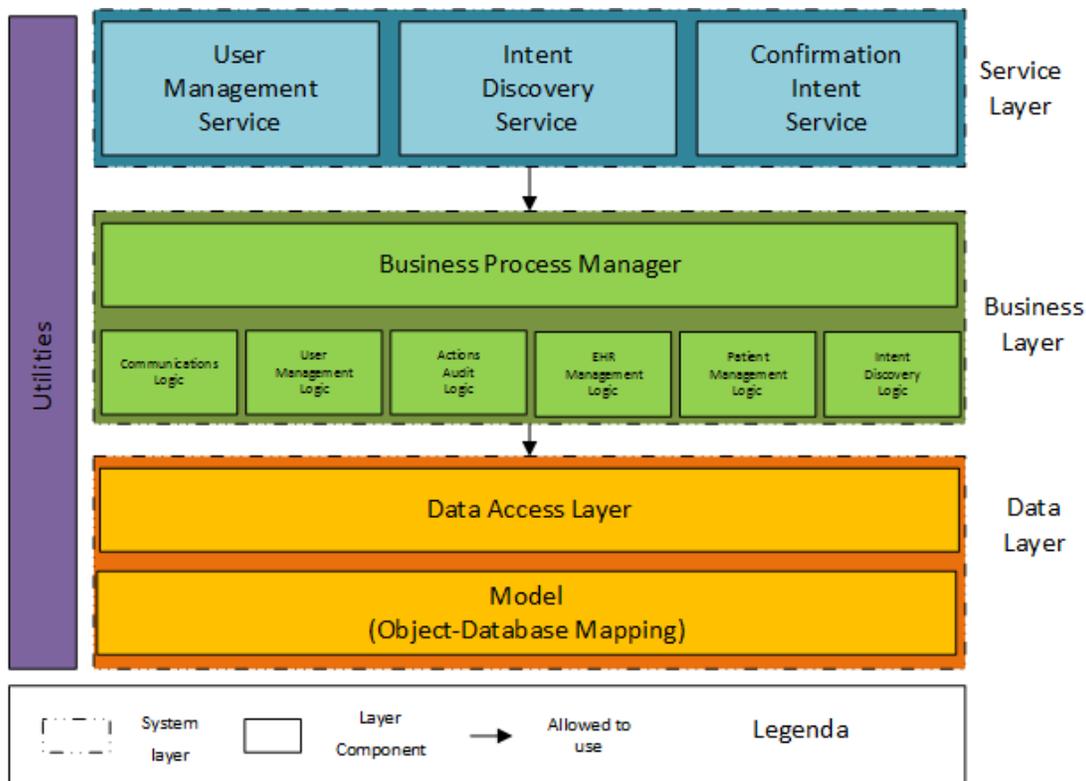


Figura 11 - Vista em Camadas do Voccallia

Service Layer Esta camada é onde se encontram todos os serviços disponíveis, no contexto do presente sistema, para serem consumidos pela aplicação móvel. Esta camada serve como interface de comunicação entre o cliente móvel e o *Voccallia*, sendo os serviços aqui presentes as diferentes forma de interação para o qual o sistema se encontra preparado lhe responder. Quaisquer outros serviços necessários serão criados nesta camada.

Business Layer Aqui encontra-se definida toda a lógica/orquestração dos processos de negócio da solução. Cada lógica de negócio é uma unidade isolada e independente das restantes, não possibilitando desta forma a relação direta com outras. Contudo, diferentes processos de negócio podem desencadear a execução de outros, sendo a sua gestão realizada a unidade hierarquicamente superior (**Business Process Manager**), como é característico da representação usada (Clements, et al., 2003).

Data Layer Esta camada diz respeito a todo o tipo de acesso e escrita de dados pelo *Voccallia*. Mais uma vez, todas as operações associadas aos dados são abstraídas por uma unidade hierárquica superior ao modelo de dados. Assim consegue-se promover a modificabilidade associada e, ainda, a testabilidade exigida no acesso às fontes de dados pela conjugação dos princípios de **IoC** e de **LSP**.

Utilities Por fim, esta camada tem a responsabilidade de prover as outras com mecanismos auxiliares de transformação de dados, abstração de acesso a recursos externos, encriptação, entre outros. Como tal, é transversal a todas as camadas, uma vez que, apesar de fornecer diferentes mecanismos para situações das diversas camadas, reforça o princípio de SRP, já que existe apenas uma camada responsável pela sua disponibilização. Assim, consegue-se prevenir a sua dispersão por toda a solução.

Por outro lado, interessa apresentar o funcionamento do sistema através da interação dos seus principais componentes, identificando mecanismos e ordem de comunicação. Como tal, Figura 12 apresenta a dinâmica de um pedido do utilizador para descoberta da sua intenção e das entidades envolvidas para confirmação de ação no sistema de informação.

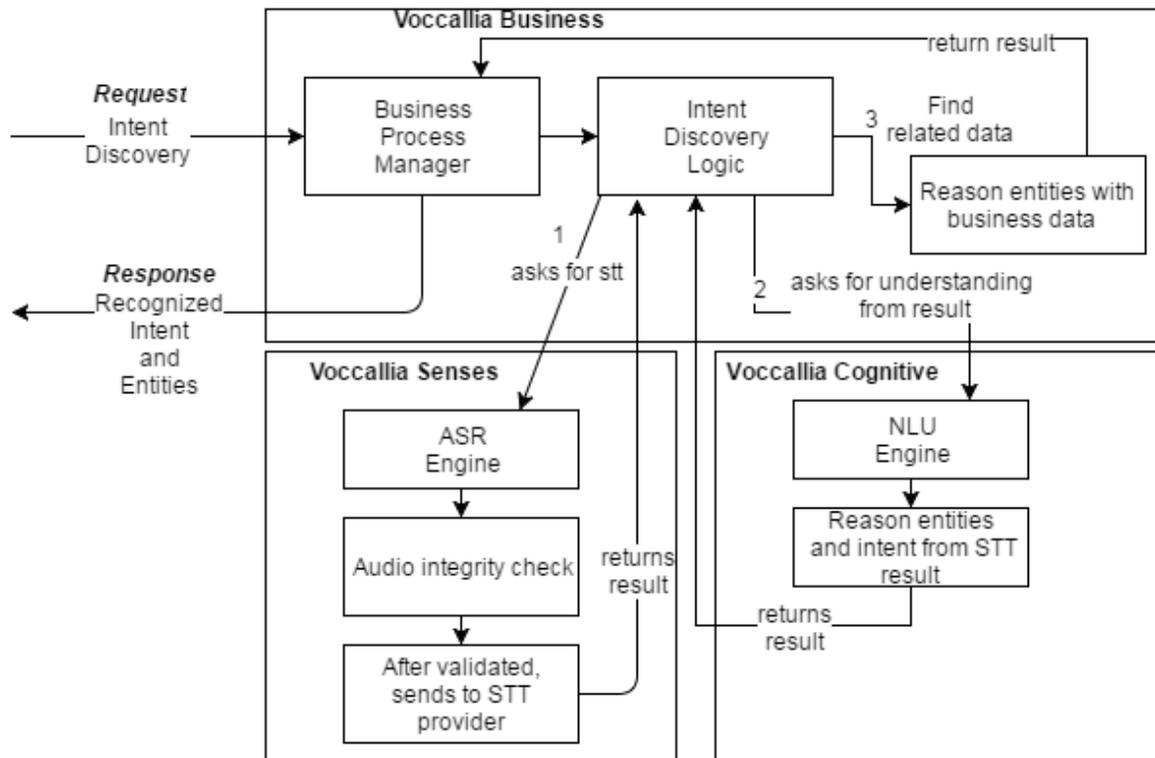


Figura 12 - Perspetiva dinâmica da arquitetura - Descoberta da intenção do utilizador a partir do áudio associado ao pedido do utilizador

Vocallia Business É responsável pela orquestração dos diversos processos de negócio e retorno da informação devidamente processada para os serviços que interagem com o sistema. Aqui é realizada toda a contextualização de entidades de acordo com o contexto do utilizador.

Vocallia Senses Este módulo tem a responsabilidade de lidar com os pedidos de “tradução” do áudio associado ao discurso do utilizador para o formato de texto. Contudo, antes sequer de

enviar o áudio para o serviço de ASR, deve realizar as operações de verificação de integridade e conversão de formato na eventualidade de o formato fornecido pelo utilizador não coincidir com o formato esperado. Apenas após a realização destas operações com sucesso, o áudio é enviado para processamento pelo serviço.

Vocallia Cognitive Este módulo tem a responsabilidade de, a partir do resultado fornecido pelo módulo *Vocallia Senses*, extrair a informação relativa à intenção do utilizador e às entidades envolvidas com recurso ao serviço de NLU. Posteriormente, é realizada a conversão do resultado para o formato esperado pelo módulo de lógica de negócio, que tratará da realização da contextualização dos resultados.

5.4.2 Análise das propriedades sistémicas e *tradeoffs*

Apesar do presente estágio se tratar do desenvolvimento de uma prova de conceito com o intuito de avaliar a fiabilidade e maturidade das tecnologias de reconhecimento de voz e processamento de linguagem natural, a posição da empresa é de que o trabalho desenvolvido possa rapidamente ser aproveitado e usado internamente para a integração com aplicações já existentes, com vista a aumentar a funcionalidade das mesmas. Por outro lado, ainda que no fim se constate que as tecnologias não são maduras o suficiente para se avançar no contexto da solução, interessa já ter um sistema base que possibilite suportar outras tecnologias que venham a surgir com o mínimo de alterações ao sistema (localizadas nos componentes associados a estas tecnologias).

Com efeito, a construção de um sistema que permita satisfazer estas duas condições tem que prever a possibilidade das tecnologias a usar no futuro virem, em última instância, a ser totalmente diferentes às estudadas (ex. diferentes fornecedores de serviços de ASR e NLU). Como tal, o sistema não pode correr o risco de se tornar dispendioso ou obsoleto pelo fato de se encontrar limitado ao uso das tecnologias estudadas e não permitir a flexibilidade de modificação ou adição de novas funcionalidades. Desta forma, uma solução capaz de reduzir futuros custos e, ainda assim, não limitar as tecnologias principais que caracterizam o sistema traz óbvios benefícios, tanto em termos da progressão/manutenção do sistema no futuro (adição de novas funcionalidades e a alteração dos componentes existentes associados às principais tecnologias encontram-se previstas na sua arquitetura) como no custo/tempo necessário para realizar qualquer mudança ao sistema, dado que a alteração de componentes ou adição de funcionalidade é uma tarefa isolada e facilmente integrável com o restante sistema.

É apresentado nas seguintes subsecções como cada um dos requisitos definidos para a solução é concretizado arquiteturalmente.

5.4.2.1 Interoperabilidade

Em termos arquiteturais, existem dois cenários de integração onde existe a necessidade de interoperabilidade: o momento em que alguém quer integrar com o *Voccallia* (ex. aplicações móveis ou outros sistemas) e o momento em que o *Voccallia* quer integrar com o EHR. Para ambos os cenários seguiu-se a mesma estratégia, que passou pela definição de uma interface de comunicação, ainda que com algumas diferenças.

Cenário de Integração: Aplicação Móvel e *Voccallia*

Em primeiro lugar, definiu-se as interfaces de integração com o *Voccallia*, ilustradas na Figura 13.

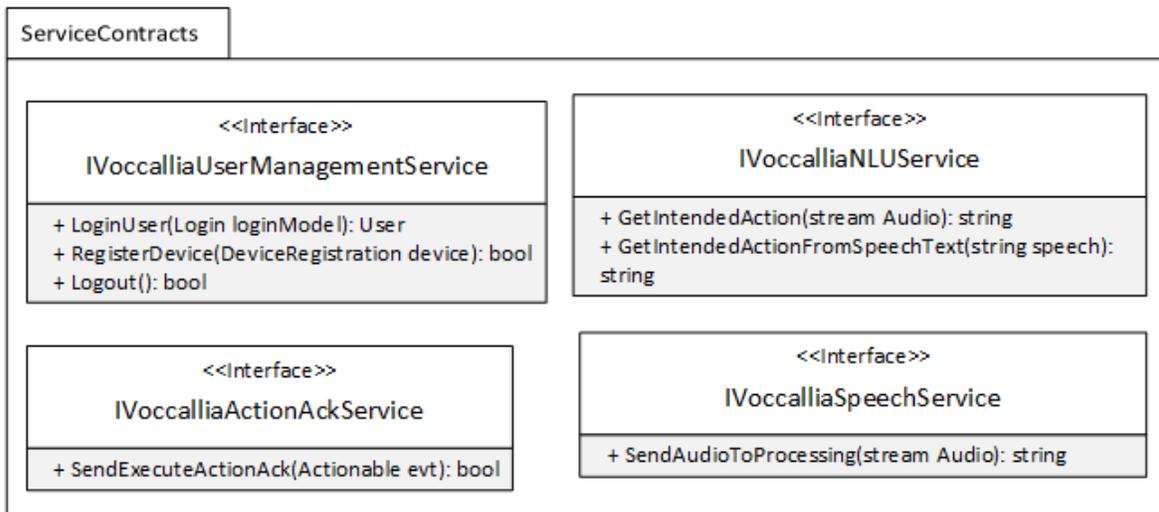


Figura 13 - Descrição dos contratos dos serviços disponibilizados para os clientes do Voccallia

Desta forma, ao se estabelecer o contrato de serviços *web* disponibilizados para os clientes móveis, consegue-se que clientes de diferentes sistemas operativos consigam eficazmente enviar pedidos para o sistema com uma mesma interface de comunicação. Contudo, de forma a poderem enviar e receber os dados, interessa a definição de um formato *standard* de resposta, agnóstico do tipo de dispositivo e/ou sistema com que se está a realizar a comunicação (possibilidade de outros sistemas/aplicações *desktop* integrarem com o *Voccallia*). Para este efeito, definiu-se o uso de um

formato de comunicação para troca de pedido-resposta entre as diferentes linguagens com uma estrutura universal, como é o caso do *Json*³⁹, cenário ilustrado na Figura 14.

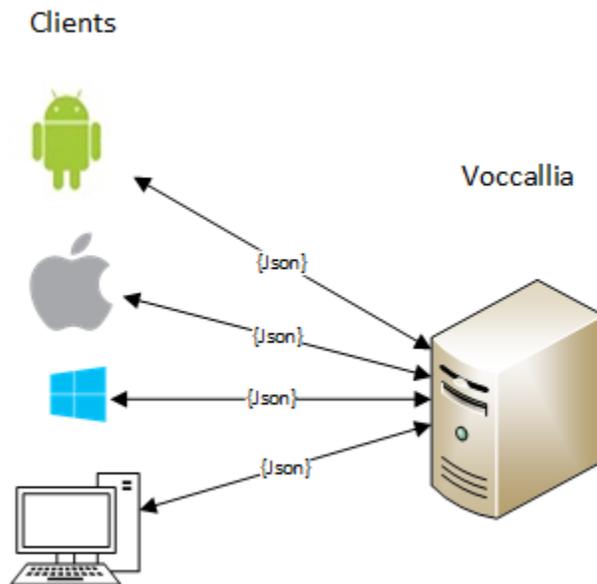


Figura 14 - Interoperabilidade entre diferentes sistemas de clientes móveis e o Voccallia

Cenário de Integração: Voccallia e EHR

Em primeiro lugar, a informação encontra-se “disponível para todos”. Isto é, com o intuito de lidar com múltiplos EHRs definiu-se uma “interface *standard*” de comunicação com o Voccallia, ilustrada na Figura 15. Esta interface define o conjunto de operações e dados que os EHRs partilham e implementam junto com o Voccallia de forma a conseguirem comunicar corretamente entre ambos.

Efetivamente, ao definir-se uma interface de comunicação entre os diferentes sistemas, sejam eles os EHRs ou outros sistemas de integração com o Voccallia, é estabelecido um contrato único de comunicação. Além disso, estas interfaces promovem a flexibilidade de alteração da informação (ex. adição ou remoção de novos campos) uma vez que não se sabe que outro tipo de informação pode ser importante trocar com entre ambos os sistemas. Por isso, é necessária a escolha de uma tecnologia de comunicação que promova estes atributos (ex. SOAP que define disponibiliza as interfaces num documento wsdI que permite a cada um dos consumidores⁴⁰ ter rápido acesso a uma estrutura partilhada e *standard* dos serviços em execução e da informação a usar num formato *standard* - XML). De realçar que, uma vez que os serviços se encontram disponíveis num formato

³⁹ <http://json.org/>

⁴⁰ Num ambiente SOA, entendem-se como consumidores as entidades interessadas em usufruir do acesso e utilização de um serviço

standard e “aberto”, qualquer sistema noutra linguagem pode usar os serviços sem limitações. Isto porque é possível a geração de *stubs* de ligação ao serviço a partir do documento de especificação

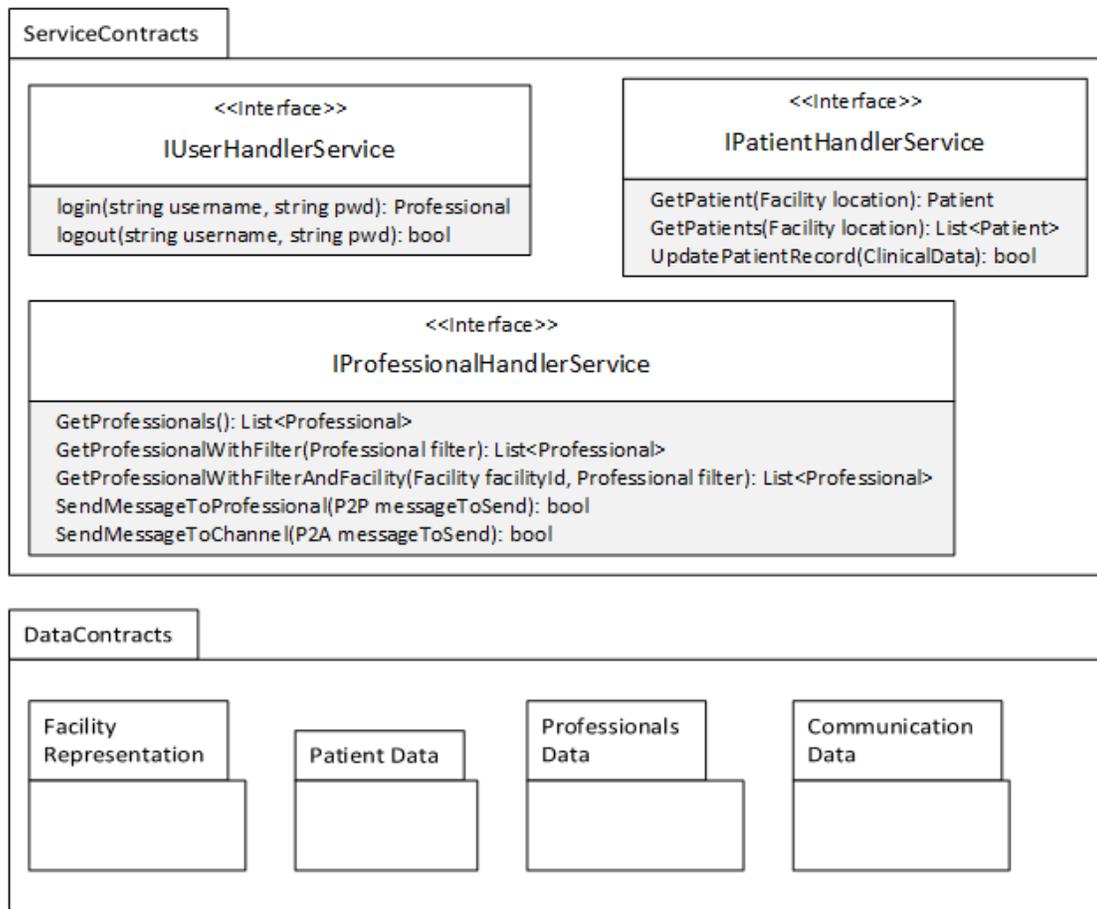


Figura 15 - Representação dos contratos de comunicação entre o conetor do EHR e o Voccallia

do serviço (WSDL) através de ferramentas já existentes (ex. no Java existem as ferramentas Apache CXF⁴¹, JAX-WS⁴²; no C# existe o *svcutil* do WCF⁴³). Isto é, não importa qual a tecnologia que o cliente use na implementação dos seus serviços, respeitando desta forma o requisito de interoperabilidade entre diferentes sistemas.

Em segundo lugar, apesar de as interfaces definidas fornecerem já uma base na garantia da interoperabilidade entre diferentes sistemas, o *Voccallia* continua a não suportar na sua arquitetura um mecanismo de autenticação do utilizador. Isto acontece porque o seu propósito é o de servir de *broker* (Broker, 2016) (Schmutz, Liebhart, & Welkenbach, 2010) entre a aplicação móvel e o EHR a

⁴¹ <http://cxf.apache.org/docs/wsdl-to-java.html>

⁴² https://docs.oracle.com/cd/E17802_01/webservices/webservices/docs/2.0/tutorial/doc/JAXWS3.html

⁴³ [https://msdn.microsoft.com/en-us/library/dd456779\(v=vs.110\).aspx](https://msdn.microsoft.com/en-us/library/dd456779(v=vs.110).aspx)

que o utilizador se deseja conectar através de um mecanismo de *Single Sign On (SSO)*. Efetivamente, **o objetivo deste sistema é mesmo o de não suportar *per se* um mecanismo autenticação**. Ou seja, **a autenticação do utilizador é**, única e exclusivamente, **da responsabilidade do EHR** ao qual o utilizador se está a querer conectar e, posteriormente, “reutilizada” pelo *Voccallia*. Contudo, tal fato não dissocia o *Voccallia* da responsabilidade de ter de ser capaz de suportar a resposta a esse mesmo mecanismo de autenticação com o intuito de permitir a interoperabilidade entre os diferentes sistemas. Desta forma, podemos verificar na Figura 17 qual a estratégia idealizada na arquitetura para lidar com este requisito.

De notar que, ao fazer uso desta estratégia, estamos também a permitir a extensibilidade no que diz respeito à adição de novos mecanismos de segurança por parte dos diferentes provedores de identidade (Ex. ADFS – *Active Directory Federation Services*). Outros benefícios do uso desta estratégia são, por exemplo, a flexibilidade dada pelo uso de diferentes mecanismos de gestão/segurança de *tokens* (ex. *OAuth*, *SAML*, etc.) que podem ser utilizados de acordo com os objetivos que se pretendam alcançar com os serviços disponibilizados (ex. autenticação, autorização ou troca de chaves de sessões); por outro lado, a esta estratégia torna a solução independente do tipo de mecanismo usado, o que por sua vez permite que diferentes protocolos de autenticação possam comunicar entre si ao fornecer uma camada superior de abstração dos protocolos existentes. Esta é uma vantagem desta estratégia quando comparada com as suas alternativas, nomeadamente o protocolo *Kerberos* e o protocolo **X.509 PKI** (Brokered Authentication: Security Token Service (STS), 2016).

Por fim, o *Voccallia*, este necessita de saber qual o endereço de cada um dos EHRs de forma a poder comunicar com eles. Torna-se então necessário a existência de mecanismos de descoberta de serviços com base no tipo de cliente que se deseja contactar. A estratégia para a descoberta dos serviços deve fazer uso de *standards* já definidos e frequentemente usados como é o caso do *WS-Discovery*⁴⁴ ainda que, para o âmbito do estágio, se tenha definido o mecanismo de descoberta dos serviços a partir de uma tabela de endereços (*service registry*), ilustrada na Figura 16, que conforme as instalações poderia ser atualizada dinamicamente.

⁴⁴ http://docs.oasis-open.org/ws-dd/discovery/1.1/os/wsdd-discovery-1.1-spec-os.html#_Toc234231798

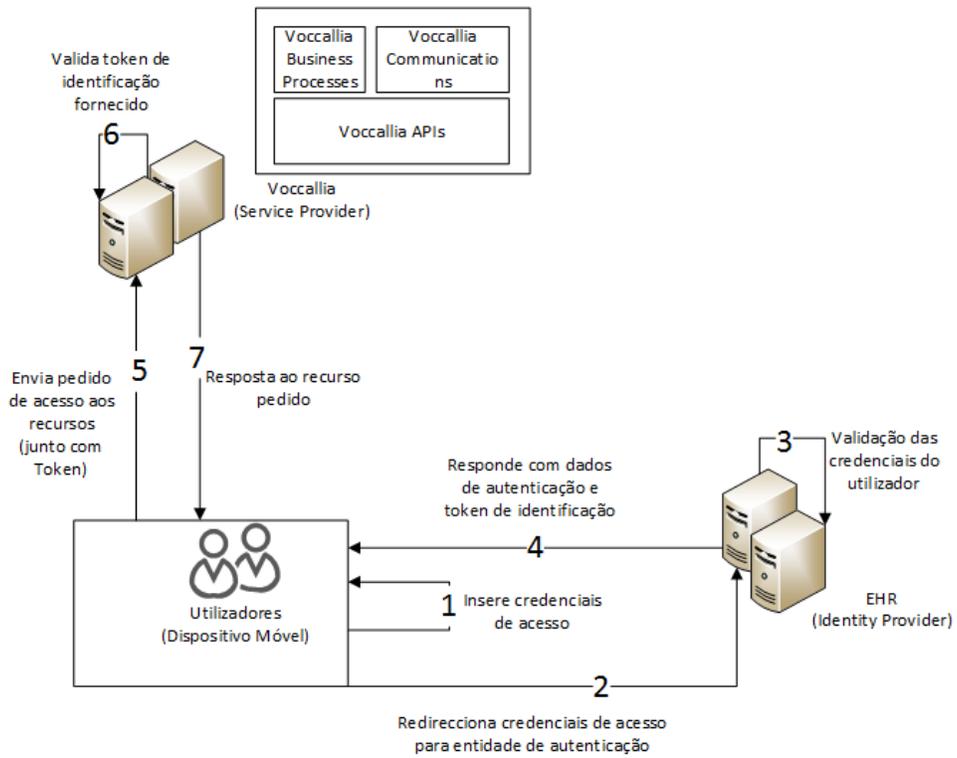


Figura 17 - Estratégia de delegação da autenticação do utilizador usada no Voccallia - Brokered authentication

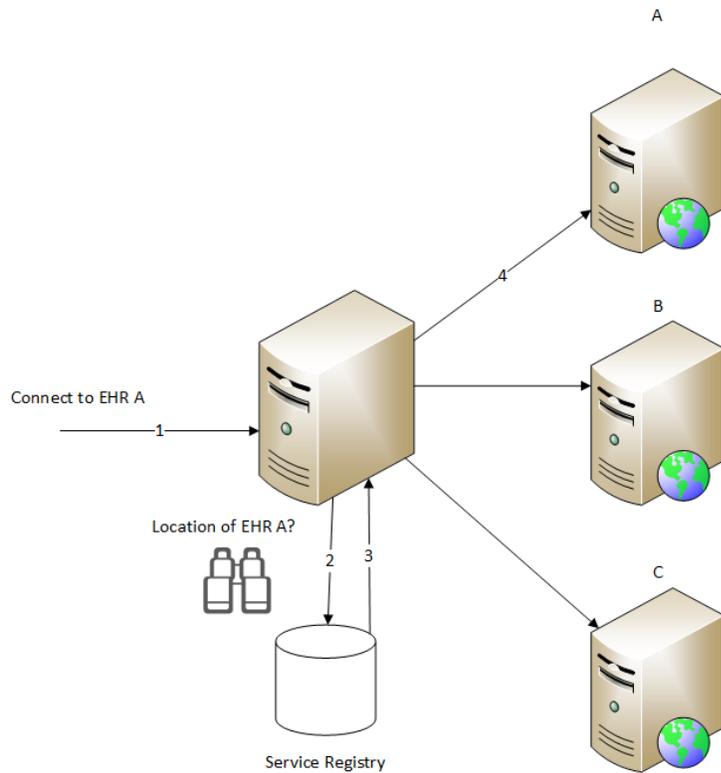


Figura 16 - Mecanismo de interoperabilidade entre o Voccallia e diferentes EHRs - EHR address lookup

5.4.2.2 Modificabilidade

Como referido na secção, o sistema deve ser capaz de acomodar um certo nível de modificações em qualquer um dos seus componentes, rapidamente e com custo reduzido. De facto, tratando-se este projeto de uma prova de conceito, é de esperar que das ferramentas utilizadas algumas delas possam não vir a fazer parte do sistema no futuro. Como tal, os módulos constituintes do sistema têm que ser idealizados logo à partida com a possibilidade de alteração dos seus componentes principais ou até mesmo lidar com a sua própria extensibilidade.

Desta forma, ao seguir os princípios arquiteturais descritos na secção 5.3 foram seguidas as seguintes estratégias arquiteturais com o intuito de lidar com os diferentes requisitos de modificabilidade existentes:

- **Aumento da coesão semântica** – que consiste na clara separação de processo (ex. processos de negócio, acesso e transformação de dados, etc.) e na definição de responsabilidades distintas e não sobrepostas para cada uma das classes e ou módulos.
- **Antecipação de alterações expectáveis** – que consiste em avaliar a atribuição de responsabilidades ao módulo através da avaliação de alterações a realizar com base num conjunto de mudanças futuras. Esta tática é utilizada em conjunto com a anterior com o objetivo de minimizar os efeitos da aplicação das alterações a um módulo.

Com efeito, no presente estágio existem três áreas principais onde tem de ser garantido este requisito. São elas os módulos de processamento cognitivo (**VocalliaCognitive**), processamento de áudio (**VocalliaSenses**) e o de orquestração e criação de atividades de negócio (**VocalliaProcessOrchestrator**). De seguida, é realizada a descrição das estratégias usadas para cada um.

VocalliaCognitive

A principal responsabilidade deste módulo é a de extrair conhecimento a partir de dados. Ou seja, este módulo é responsável pelo processamento do texto resultante do *speech-to-text* do módulo *VocalliaSenses* através de mecanismos de *spell-checking* e extração de entidades⁴⁵ resultantes do processamento semântico fornecido pelo serviço de NLU. Todavia, admitindo a necessidade de futuramente se querer alterar o serviço de NLU ou de querer dar suporte a outros torna-se

⁴⁵ Para mais detalhes, consultar o capítulo 3 com o objetivo de perceber as entidades resultantes do processo de NLU.

imperativa a capacidade de o sistema conseguir lidar com tais modificações com um mínimo de alterações e perda de disponibilidade junto com os outros módulos. Como tal, definiu-se logo que a arquitetura deste módulo deveria funcionar com base numa arquitetura de *plugins* com o intuito de minimizar as dependências com outros módulos e abstrair a implementação de cada um dos serviços de NLU a partir de uma interface comum de métodos e dados, como se pode verificar na Figura 18.

Adicionalmente, ao se aplicar aqui o princípio de *IoC* (definido *secção 5.3*) torna-se possível tornar o processamento semântico independente do serviço que o está a suportar.

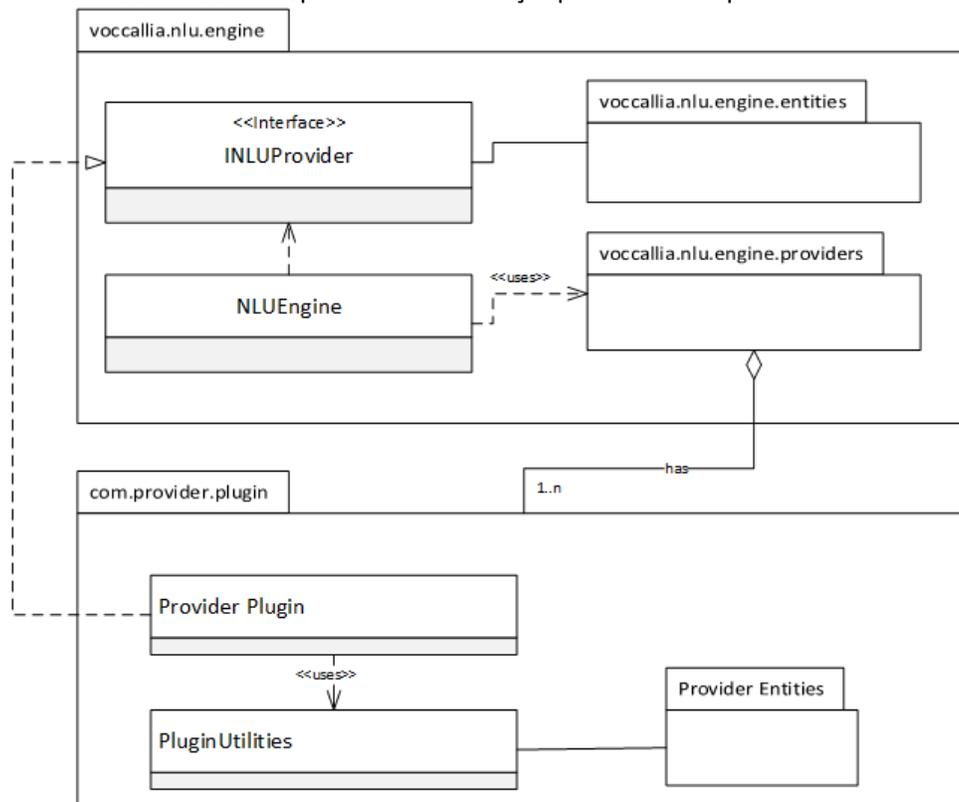


Figura 18 - Solução arquitetural de plugins para o módulo VoccalliaCognitive

VoccalliaSenses

A principal responsabilidade deste módulo é a de abstrair todo o processamento relativo ao áudio enviado pelo utilizador (ex. transformações do áudio, envio para o serviço de STT, etc.). Assim, e sendo este módulo semelhante ao anterior em termos de requisitos (modificabilidade e extensibilidade), foram utilizadas as mesmas estratégias de abstração e de arquitetura – arquitetura

baseada em *plugins* e com inversão de dependências (*IoC*). A Figura 19 apresenta o esquema da arquitetura do módulo.

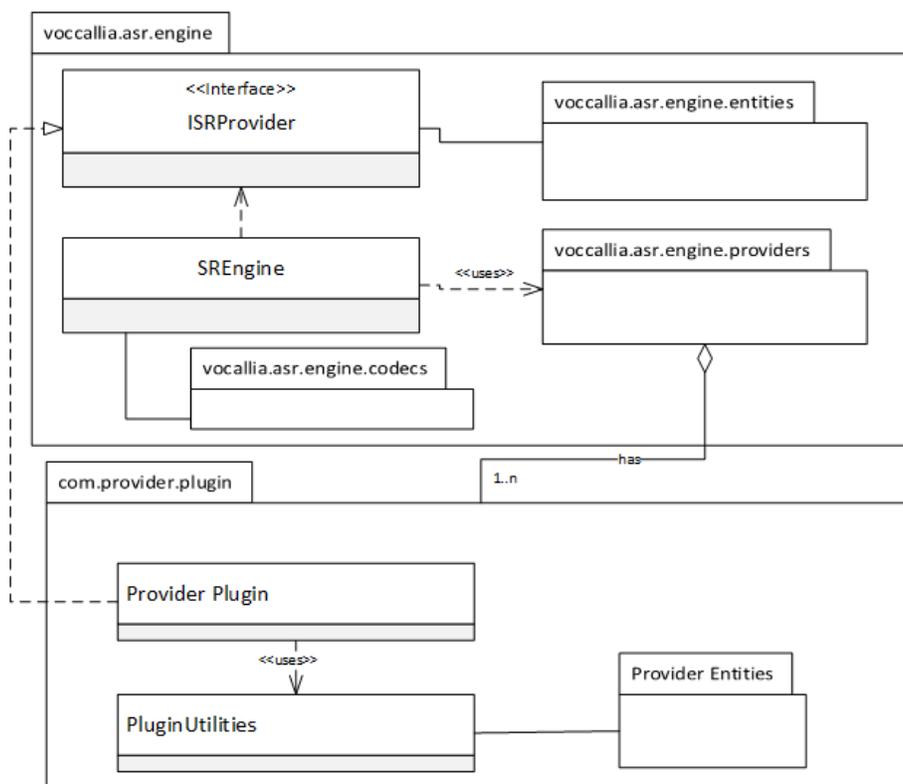


Figura 19 - Solução arquitetural de plugins para o módulo VoccalliaSenses

VoccalliaProcessOrchestrator

Neste módulo garante-se o requisito de modificabilidade em diferentes níveis através do uso de diferentes táticas arquiteturais.

Em primeiro lugar, e sendo este módulo aquele que inicia todo o processo de comunicação com os restantes intervenientes do *Voccallia* é necessária a existência de um ponto de acesso à camada de dados própria do sistema. Para este efeito, criou-se uma abstração de acesso a dados por intermediário, mais concretamente, seguindo o *design pattern* de repositórios, ilustrado na Figura 20. O acesso a dados é, então, realizado a partir de um repositório por este módulo.

Este tipo de abstração é bastante útil uma vez que, para além de satisfazer o requisito de modificabilidade, satisfaz o requisito de testabilidade (The Repository Pattern, 2016).

Por sua vez, é também neste módulo que os processos de negócio são criados para darem seguimento a realização das ações do utilizador no sistema. Uma vez que os processos de negócio

podem ser modificados e adicionados conforme as necessidades de negócio do produto e da empresa, considerou-se que também definir uma “arquitetura de *plugins*” para dar permitir a

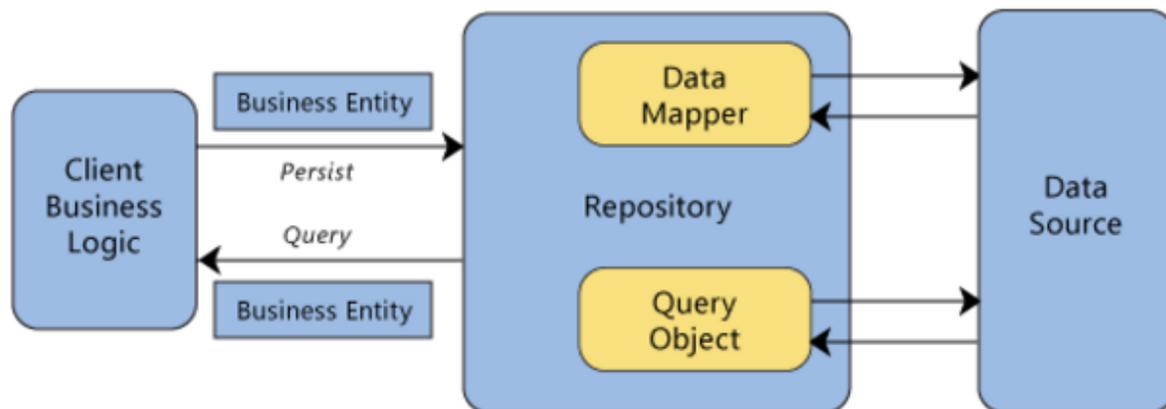


Figura 20 - Design Pattern de Repositório usado para abstrair o acesso aos dados, imagem (The Repository Pattern, 2016)

adição ou alteração de processos de negócio em *runtime* e de acordo com os processos de execução característicos de cada EHR. Isto porque, apesar de os processos de negócio serem exteriores ao Voccallia, a sua execução continua a ser despoletada por processos internos com a informação respetiva. Deste modo, diferentes EHR podem requerer que de um processo de negócio do Voccallia possa ser despoletado um ou mais processos seus (ex. Alta de paciente, para além de envolver a comunicação com os responsáveis médicos, pode envolver a chamada aos processos de faturação). A solução arquitetural para este módulo é visível Figura 21 de onde se pode ver como o processo de modificação e adição dos processos de negócio é gerido por uma entidade de mediação entre as implementações definidas de base e as adicionadas posteriormente. A injeção do tipo implementação de processos de negócio a usar pode ser realizada via configuração (ex. ficheiro xml).

Para concluir, importa destacar o fato da presente solução arquitetural permitir a injeção de processos de negócio independentemente da tecnologia que possa vir a ser utilizada na sua definição. Isto porque, ao fazer uso dos princípios de IoC, DI e ISP, a definição dos processos de negócio deixa de ser uma especificação concreta definida à partida para todos. Como se pode ver na Figura 21, existem dois tipos de processos de negócio que podem ser injetados para a entidade de mediação (os *BaseBps* e os *CustomBps*). Os primeiros dizem respeito aos processos base fornecidos pelo Voccallia. Por seu lado, os restantes referem-se aos processos especificados pelas entidades que integram com o Voccallia. Desta forma, os processos de negócio podem ser criados

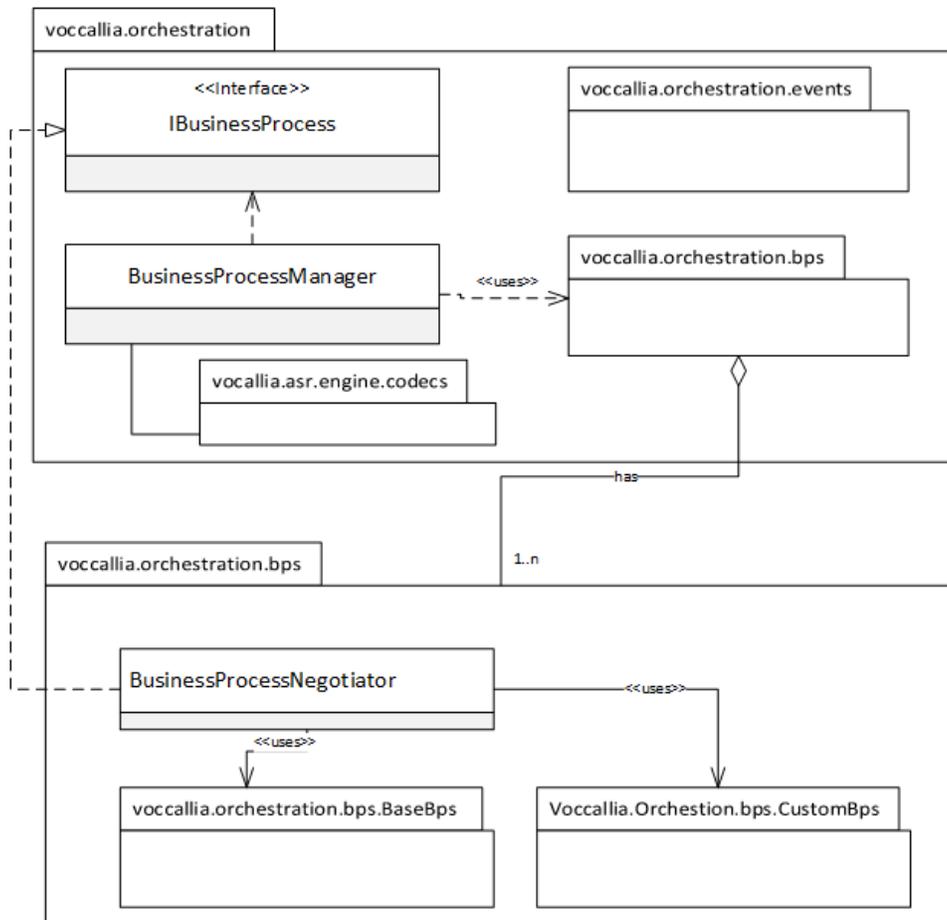


Figura 21 - Solução arquitetural definida para o módulo VocalliaOrchestration

de acordo com as necessidades de cada empresa, tornando-se assim uma solução genérica que se pode adaptar a qualquer cenário de definição de processos e tecnologias.

5.4.2.3 Testabilidade

Por fim, o requisito de testabilidade é concretizado através da especialização de interfaces de teste com o intuito de controlar e observar o estado interno dos componentes do sistema. Por outro lado, a especificação da arquitetura do sistema com base nos princípios arquiteturais especificados na secção 5.3, nomeadamente, o princípio *SOLID* dadas as características por ele promovidas. Quero com isto dizer que, sendo este princípio claramente direcionado para a clara separação de responsabilidades entre componentes (classes no seu nível mais baixo), todo o seu *design* é inerentemente testável uma vez que com a possibilidade de as dependências existentes poderem ser facilmente substituídas ou removidas, é promovido o isolamento do teste de classes e métodos (princípio de SRP e ISP); por outro lado, a capacidade de abstrair ou injetar uma dependência possibilita que os diferentes componentes possam ser substituídos independentemente (LSP e DIP).

6 Trabalho Desenvolvido

Neste capítulo é demonstrado o trabalho desenvolvido no durante o presente estágio. De notar que o trabalho desenvolvido compreendeu a realização de diferentes atividades para além da solução alvo. Isto porque, além de se pretender a criar uma solução com as capacidades de suporte às tecnologias de ASR e NLU, procurou-se estudar a viabilidade e robustez das ferramentas estudadas para cada uma das áreas através da criação de ferramentas adicionais de ajuda à recolha de amostras de áudio e à realização dos respetivos testes de “performance” às ferramentas de ASR. Por outro lado, e dada a natureza interativa e de integração com outros sistemas da solução desenhada, procurou-se criar os mecanismos adicionais de prova do conceito (ex. Conetor *MedicineOne* e aplicação móvel – *Android*).

6.1 Desenvolvimento da Solução

O desenvolvimento da solução caracterizou-se pela implementação dos mecanismos de suporte a seguir descritos.

Modelo de dados

O modelo de dados para a solução é bastante simples. Isto acontece porque a solução pretende integrar com diferentes sistemas de informação externos sendo apenas da responsabilidade do Voccallia a gestão dos utilizadores e dos respetivos EHRs. Na verdade, muito pouca informação tem de ser persistida no Voccallia. Os únicos dados persistidos, para além da informação do utilizador e dos EHRs, são as comunicações realizadas entre utilizadores. Assim, na Figura 22 é apresentado o modelo de dados conceptual que caracteriza o sistema.

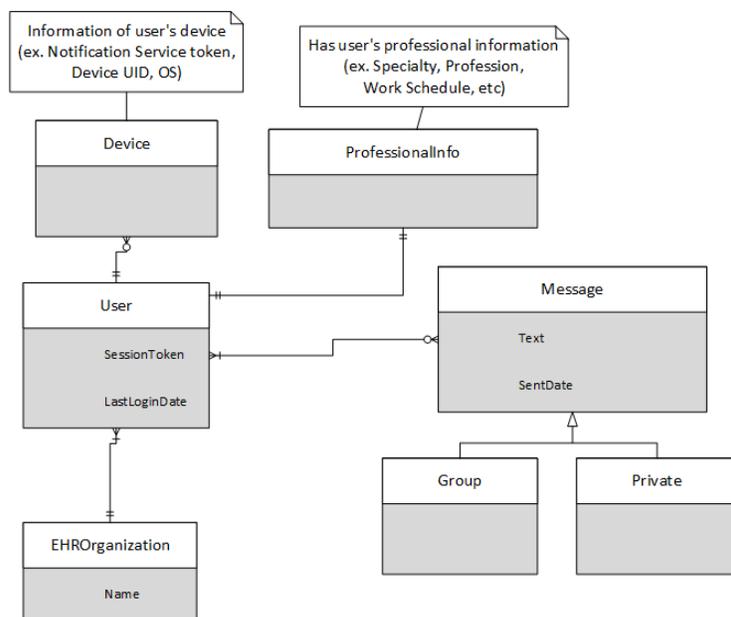


Figura 22 - Modelo conceptual de dados do Vocallia

Extensibilidade / Modificabilidade dos componentes do sistema

Como já foi apresentado e descrito na secção 5.4, foram implementadas diferentes estratégias e princípios com o intuito lidar com os requisitos de modificabilidade e extensibilidade no desenvolvimento da arquitetura. De fato, relembra-se que toda a arquitetura assenta nos princípios S.O.L.I.D.⁴⁶ de forma a garantir uma maior longevidade do sistema. Apesar destes princípios se estenderem às práticas de desenvolvimento do sistema (*design patterns*), importa identificar que outras tecnologias foram utilizadas para a sua concretização. É neste contexto que surge a tecnologia *Managed Extensibility Framework* (MEF).

O MEF é uma biblioteca integrada no .NET Framework a partir da versão 4.0 com o objetivo de permitir a uma determinada aplicação ser extensível. É baseada nos princípios de IoC e de *Dependency Injection* (DI) e distingue-se das *frameworks* existentes por ser capaz de gerir um conjunto de **dependências desconhecidas pelo sistema/aplicação**. Ou seja, novas dependências podem ser injetadas num momento posterior ao da compilação. Isto é conseguido pela capacidade do MEF em possibilitar a descoberta em *runtime* de componentes (no contexto, designado de

⁴⁶ Ver secção 5.3

partes) disponíveis através da tarefa de composição⁴⁷. O processo de descoberta das dependências é ilustrado na Figura 23.

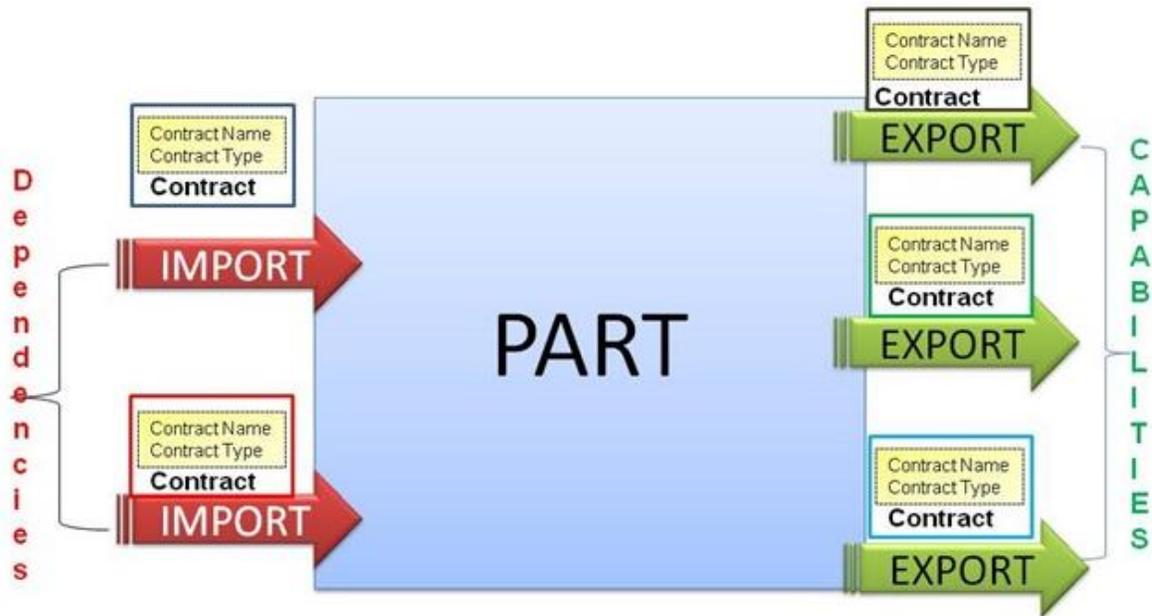


Figura 23 - Composição e descoberta de dependências com o MEF, imagem adaptada (htt)

Como se pode verificar, a composição das dependências é realizada a partir dos contratos que caracterizam as partes. No caso do Voccallia, a composição das dependências foi realizada a partir dos contratos de especificação das ações referentes aos fornecedores (em inglês, *providers*)⁴⁸ das funcionalidades de ASR e NLU (*IASRProvider* e *INLUProvider*) e, também, do tipo de processos de negócio que podem ser usados no sistema (*IBusinessProcess*).

Tradução de Ações em Processos de Negócio

Uma das características mais proeminentes do Voccallia é a realização de ações de negócio de acordo com a intenção expressa pelo utilizador. Isto é, as ações que um utilizador envia a partir dos comandos de voz para o Voccallia devem refletir-se na execução dos processos de negócio correspondentes. Isto é conseguido através das seguintes estratégias de implementação:

1. Descoberta da intenção do utilizador a partir de um conjunto pré-definido de ações criadas no serviço de NLU.
2. Correspondência da intenção reconhecida com o tipo de processo de negócio.

⁴⁷ [https://msdn.microsoft.com/en-us/library/dd460648\(v=vs.110\).aspx](https://msdn.microsoft.com/en-us/library/dd460648(v=vs.110).aspx)

⁴⁸ Implementações concretas em torno de cada serviço (ex. *providers* de ASR – Nuance e Google)

3. Envio do comando de execução do processo de negócio no EHR.

É evidente que o processo descrito em cima é sempre posterior à aprovação do utilizador sobre a informação respetiva à descoberta de intenção e restantes dados envolvidos (ex. entidades envolvidas, dados a inserir, etc) sendo, por isso, o tipo de ação de negócio “dependente” do aval do utilizador. Isto porque, apesar da ótima experiência de utilização que o serviço de NLU nos tem proporcionado, a total dependência do Voccallia no serviço para a correta descoberta da intenção do utilizador poderia causar erros ou falhas na execução dos respetivos processos de negócio (ex. um médico deseja que seja efetuado o transporte do utente a outro serviço → a ação de negócio a executar seria **Transporte de Utentes**; contudo, se o serviço reconhecesse outra intenção – como por exemplo, **Alta de Utente** – a ação de negócio já seria **Processo de Alta de Utente**).

O mecanismo de tradução é baseado no *design pattern* de comandos (em inglês Command Pattern (Gamma, Helm, Johnson, & Vlissides, 1994).

Suporte de NLU

Face aos requisitos de NLU já identificados na secção 3.3, a escolha da ferramenta para o presente estágio recaiu pela escolha da ferramenta api.ai em detrimento da wit.ai. As razões da escolha foram, comparativamente à wit.ai, baseadas na maior confiança e robustez associadas ao correto reconhecimento das entidades e intenções presentes no discurso do utilizador. Aliadas à facilidade e simplicidade de configuração dos cenários de utilização e criação das entidades envolvidas, disponibilizadas através de uma plataforma *web*, torna-se uma solução bastante poderosa para o suporte à vertente de NLU.

Efetivamente, toda a configuração dos cenários foi bastante rápida e efetiva uma vez que todo o processo é suportado por uma vasta documentação, sendo muitas vezes acompanhada de exemplos de cenários de uso com os passos todos necessários de preparação da ferramenta. De realçar, também, a qualidade e disponibilidade da equipa de suporte da empresa que muitas vezes forneceu explicou e esclareceu dúvidas relacionadas com a implementação e desenvolvimento dos cenários de utilização do projeto.

Assim, no Anexo A, são apresentadas as configurações realizadas na plataforma para o correto reconhecimento das entidades e intenções necessárias para satisfazer os cenários de utilização descritos pelos requisitos funcionais apresentados na secção 4.2.

Como foi referido anteriormente, o sistema Voccallia tem acesso à plataforma api.ai através da implementação de um conetor que comunica com a API HTTP disponibilizada pela plataforma, como foi identificado na secção 3.3. A sua gestão é realizada pelo módulo VoccalliaCognitive.

6.2 Desenvolvimento dos Sistemas de Integração

Nesta secção é apresentado o trabalho adicional realizado com vista a exemplificar a integração *end-to-end* da solução desenvolvida. Para este efeito, foram realizadas integrações do sistema tanto ao nível business-to-business (*B2B*) como *business-to-consumer* (*B2C*) com vista a ilustrar o cenário geral de funcionamento do *Voccallia*. A integração B2B foi realizada com o desenvolvimento de um conetor de acesso ao EHR MedicineOne (subsecção 6.2.1) e a B2C através do desenvolvimento de uma aplicação móvel de interface com *Voccallia* (subsecção 6.2.2).

6.2.1 Desenvolvimento do Conetor *MedicineOne*

Quando se cria uma solução cujo objetivo é o de permitir a interoperabilidade com outras entidades é necessária a criação de mecanismos de integração entre o sistema (*Voccallia*) e a outra entidade. Este tipo de integração é realizada com o intuito de expor funcionalidades e/ou processos de negócio da empresa para terceiros, como parceiros de negócio, fornecedores e clientes (Schmutz, Liebhart, & Welkenbach, 2010).

Assim, como já foi previamente mencionado na secção 5.4.2.1, a integração B2B foi realizada através da definição de contratos de partilha de funções de negócio previamente acordados entre ambas as entidades. Em termos práticos, a implementação dos serviços especificados para o conetor (neste caso, o *MedicineOne*) expõem os diferentes domínios de negócio internos necessários para a realização do processo de negócio do *Voccallia*. Assim, na Figura 24 pode-se ilustrar qual a perspectiva de integração realizada com o desenvolvimento do conetor.

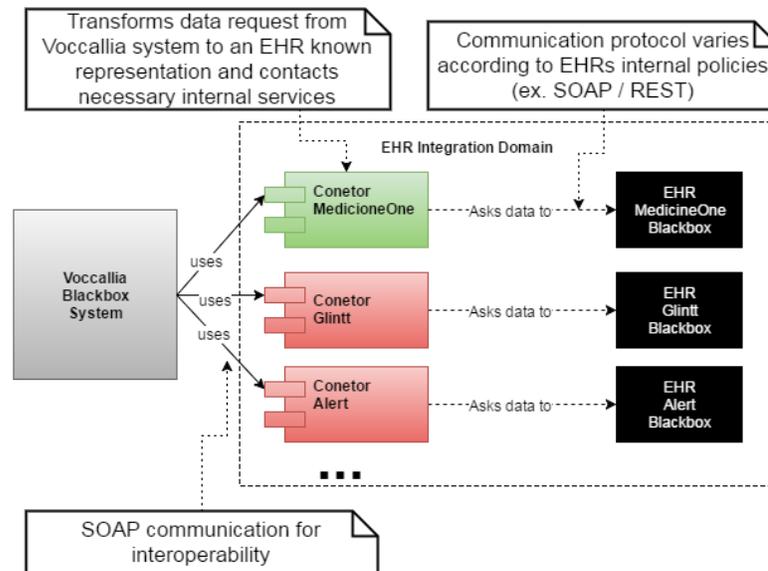


Figura 24 - Domínio de integração do conetor

6.2.2 Desenvolvimento da aplicação móvel

O objetivo do desenvolvimento da aplicação móvel foi o de validação de prova de conceito do sistema e tecnologias adjacentes aos cenários de utilização aqui desenvolvidos. De realçar que a este nível não foram especificados quaisquer requisitos não-funcionais dado que o desenvolvimento da própria aplicação nem sequer estar previsto no planeamento inicial de trabalhos do presente estágio e servir apenas de suporte à prova de conceito. Todavia, associado ao fato de uma aplicação móvel ajudar a compreender os mecanismos de interação do utilizador e necessidades de um contexto real e, também, ajudar a proporcionar um artefacto mais palpável para os principais *stakeholders* do sistema, o seu desenvolvimento foi reconsiderado para o período referente ao estágio.

Deste modo, são ilustrados na Figura 25 *screenshots* representativos dos principais ecrãs da aplicação com o propósito de ilustrar as diferentes fases que caracterizam o processo de interação com o sistema *Voccallia* e como estas têm impacto na agilização de processos dos profissionais de saúde. Aqui podemos dar mais destaque ao segundo e terceiro ecrãs que representam os principais menus de interação dos utilizadores com a aplicação. No segundo, o utilizador carrega no botão de forma a gravar a ordem ou pedido que deseja realizar. Já no terceiro temos a resposta da intenção descoberta pelo sistema para o áudio gravado pelo utilizador. Neste caso, é expressa a intenção de definição de contexto de um paciente.

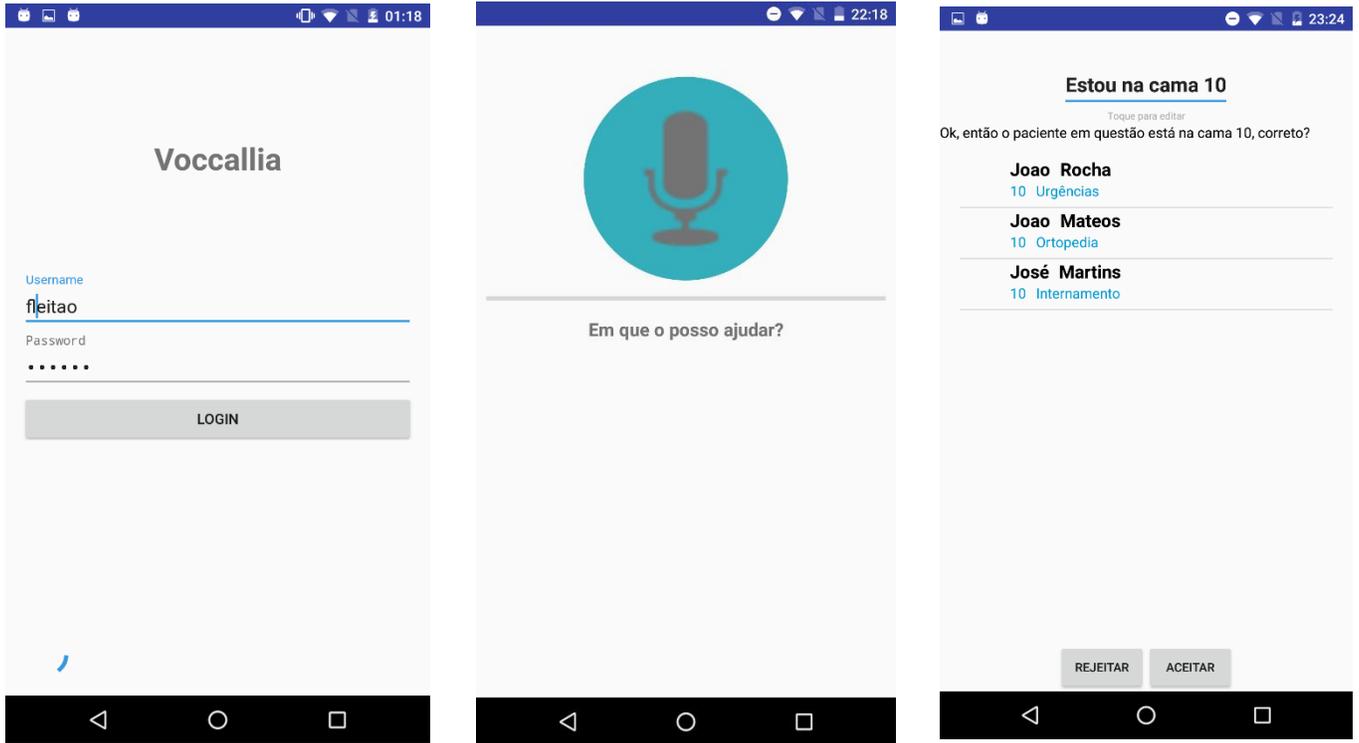


Figura 25 - Principais menus da aplicação desenvolvida

7 Testes

Este capítulo apresenta os testes realizados no presente estágio de forma a assegurar a realização das funcionalidades e requisitos definidos. Por outro lado, dado o objetivo principal do presente trabalho se focar na avaliação das ferramentas de ASR, são também apresentados os testes e ferramentas auxiliares usadas no processo de *benchmarking*.

7.1 Testes de aceitação

O principal objetivo deste tipo de testes é a validação da realização dos requisitos funcionais de acordo com o comportamento e resultados esperados pelo utilizador final. A realização destes testes permitiu uma melhor definição do comportamento de interação que deve caracterizar o funcionamento da aplicação móvel e, ainda que parcialmente, o do sistema desenvolvido. Isto porque, de forma a suportar algum do comportamento esperado da aplicação móvel, alguns comportamentos do sistema tiveram que ser redefinidos.

Os testes de aceitação realizados encontram-se disponíveis no Anexo C.

7.2 Testes de Integração

Este tipo de testes é essencial no desenvolvimento de sistemas onde existem necessidades de interoperabilidade e integração de outros sistemas e/ou componentes. Aqui são realizados testes de incompatibilidades e de interfaces entre componentes de forma a avaliar o seu correto funcionamento. Assim, tratando-se o Voccallia de um sistema em que se pretende a integração de múltiplos componentes, impõe-se a realização destes testes de forma a garantir que a adição de novas peças ou a modificação das já existentes não introduz incompatibilidades (Ammann & Offutt, 2008).

Deste modo, apesar de num contexto de produto ser necessária a realização dos testes a todos os componentes de integração, no âmbito do presente trabalho apenas foram definidos testes aos principais componentes onde se antecipam maior necessidade de modificabilidade e extensibilidade. São eles os componentes de ASR (módulo VoccalliaSenses), orquestração de processos de negócio (módulo VoccalliaOrchestrator), de NLU (módulo VoccalliaCognitive) e gestão de dados (módulo VoccalliaDAL).

7.3 Plataforma de recolha de amostras de áudio para teste

O desenvolvimento desta plataforma decorreu da necessidade de se avaliar a viabilidade e robustez das ferramentas existentes de reconhecimento de voz a fim de se usar no desenvolvimento de aplicações do universo MedicineOne. Desta forma, este estudo pressupõe que a necessidade de um conjunto de amostras de áudio para que depois se possam realizar os testes às ferramentas selecionadas (ver secção 7.2).

Para este efeito, houve então a necessidade de criação de uma plataforma com mecanismos de autenticação com o intuito de limitar o acesso ao círculo de pessoas da empresa e, ainda, com capacidades para prover aos utilizadores mecanismos de gravação de voz e possibilitar aos utilizadores uma experiência próxima daquele que é o cenário real de utilização. Desta forma, no seu desenvolvimento teve-se em consideração os seguintes aspetos:

- Os cenários de utilização são no âmbito da utilização de dispositivos móveis.
- Os utilizadores não conhecem o âmbito de utilização da plataforma e necessitam de formação.
- A recolha de áudio deve abranger todos os casos de uso previstos para o presente estágio.
- A utilização da plataforma deve agilizar o processo de preparação das amostras para o estágio de testes.
- Existem diferentes OS para dispositivos móveis.
- O tempo para o desenvolvimento da plataforma é limitado (finalizada no 1º semestre).

Assim, a partir dos aspetos mencionados decidiu-se, internamente, que a plataforma desenvolvida deveria ter em mente a utilização multiplataforma com suporte às tecnologias de gravação de áudio salvaguardando que o tempo necessário para a sua criação deveria ser reduzido. Efetivamente, dados estes requisitos, o formato *web* para o desenvolvimento da plataforma, ilustrada na Figura 26, surge como sendo o mais indicado uma vez que, por si só, garante a compatibilidade de execução com os diferentes OS existentes e é mais ágil o desenvolvimento dada a familiarização com a maior parte das tecnologias de desenvolvimento (*HTML5*, *AngularJS*, **WebRTC**⁴⁹, entre outras) em comparação com o desenvolvimento individual de aplicações de suporte a cada um dos OS (desde os sistemas *desktop* ao sistemas para dispositivos móveis). Por outro lado, com o intuito de avaliar os diferentes cenários de utilização na recolha das amostras de áudio e de preparar os utilizadores para o uso da plataforma foi realizada a especificação de um total de cinquenta e duas frases



Figura 26 - Plataforma web desenvolvida para a recolha de amostras de áudio

(cenários de utilização) divididas por três categorias (cenários de comunicação, Leitura de biometrias e Relato de eventos), juntamente com a realização de uma sessão de apresentação e esclarecimento de dúvidas sobre os mecanismos de interação sendo, posteriormente, demonstrada a gravação de alguns casos de utilização.

⁴⁹ <https://webrtc.org>

Com o objetivo de agilizar a recolha de amostras e de promover a rapidez do processo foram criados dois mecanismos para realizar a recolha do áudio: modo de geração aleatória de cenários (na Figura 27) e o modo de geração normal (na Figura 28). Na verdade, a criação destes mecanismos deveu-se ao fato dos participantes da recolha serem colegas de trabalho que, devido à carga horária e, ainda,



Figura 27 - Modo de geração aleatória



Figura 28 - Modo de geração normal

às tarefas pessoais, muitas vezes queriam um processo menos interativo com a plataforma com vista a possibilitar a realização de outras tarefas aquando da realização da gravação.

O desenvolvimento desta ferramenta contemplou ainda as tarefas de limpeza e conversão de áudio para os diferentes formatos suportados por cada uma das ferramentas de ASR, nomeadamente, FLAC para a da Google e WAV para a Nuance (com a redução da taxa de amostragem para 22KHz⁵⁰). A tarefa de limpeza do áudio consistiu na remoção de “lixo”⁵¹ existente no áudio gravado a partir de dispositivos móveis, mais concretamente em dispositivos Android. Supõe-se que este problema seja adjacente da biblioteca usada para o processo de gravação.

Por sua vez, a tarefa de conversão do áudio foi necessária, como já foi mencionado, para dar suporte às diferentes ferramentas de ASR e compreendeu o uso de ferramentas adicionais com o intuito de transformar o áudio gravado (formato WAV com taxa de amostragem de 44.1 KHz e codificação com 16 bits) nos formatos por elas suportado. Contudo, apesar de esta tarefa ter sido realizada com recurso a ferramentas de tratamento de áudio conhecidas, como por exemplo o NAudio⁵² e o FFmpeg⁵³, não é assegurado que o processo foi devidamente conduzido dada a inexperiência do estagiário nesta área de processamento. Isto, por sua vez, pode resultar na introdução de ruído e deformação no áudio.

Com o intuito de possibilitar uma eventual correção do processo e a realização de futuros testes sobre as mesmas amostras de áudio, o processo de transformação do áudio é realizado sempre em memória a partir de cópias destas amostras. Desta forma, evita-se a propagação de erros adicionais.

Pontos positivos

Efetivamente, o desenvolvimento desta plataforma permitiu uma forma rápida e simples recolher um conjunto vasto e diverso de amostras de áudio para os cenários de utilização que permitiu avaliar a realização de testes às ferramentas de reconhecimento de voz, podendo avaliar o comportamento das ferramentas a partir de um conjunto mais rico de parâmetros. De fato, dada a multiculturalidade das pessoas presentes na empresa, pode-se avaliar o impacto que parâmetros como o timbre da voz (grave / agudo), a entoação das palavras (caraterístico do efeito regional) e, também, o género têm no correto reconhecimento das frases/cenários especificados.

⁵⁰ Maior taxa de amostragem suportada pela ferramenta

⁵¹ Sequência de bytes extra adicionados no áudio

⁵² <https://naudio.codeplex.com/>

⁵³ <https://ffmpeg.org/>

Por outro lado, o fato de se ter pensado desde logo categorizar os cenários de utilização e fazer associação do áudio gravado respetivo junto com a informação relevante do utilizador, permitiu que uma maior estruturação dos dados para a posterior realização dos testes e respetiva análise.

Pontos Negativos

Apesar dos cenários de utilização definidos nos permitirem ter uma noção da naturalidade com que os utilizadores conseguem, efetivamente, expressar-se junto dos dispositivos móveis para a realização das tarefas, eles não se encontram expressos nem forma nem com o conteúdo com que os reais intervenientes, possivelmente, se expressariam. Quero com isto dizer que, muito provavelmente, os profissionais que atuam no contexto para o qual a solução se dirige têm formas comuns de expressão para os diferentes cenários de utilização que não se encontram previstas no conjunto especificado. Desta forma, não se consegue avaliar como determinadas expressões e determinadas palavras seriam reconhecidas pelas ferramentas de ASR nem quais os fatores que poderiam influenciar o seu reconhecimento.

Por outro lado, o fato da recolha da maioria das amostras ter sido realizado em ambiente controlado (no local de trabalho – MedicineOne - em que o ambiente é pouco ruidoso) e, obviamente, oposto ao ambiente que caracteriza, por exemplo, uma área de internamento hospitalar. Como tal, isto impede a “correta avaliação” dos testes de performance realizados às ferramentas de ASR uma vez que as condições de realização dos testes e do ambiente alvo de uso da solução não condizem.

7.4 Testes às ferramentas de ASR

Como foi apontado logo no início deste documento, na secção 1.1, o presente estágio tem como objetivo a avaliação da maturidade das tecnologias de ASR e de NLU com o objetivo de verificar a possibilidade de investimento futuro no desenvolvimento de uma solução de mercado que possa facultar diretamente aos utilizadores a possibilidade de interação através de uma aplicação móvel ou, por outro lado, permitir a aplicações e sistemas já existentes integrar com o sistema desenvolvido a fim de disponibilizar essas mesmas funcionalidades.

Assim, a avaliação da performance das ferramentas ASR foi realizada a partir da métrica de *Word Error Rate (WER)* frequentemente utilizada na realização de *benchmarks* neste âmbito (ELRA, 2016) (Gaida, et al., 2014) (Morbin, et al., 2013).

Dados Utilizados

A partir da ferramenta descrita na secção anterior foram recolhidas as gravações de voz que serviram de teste às ferramentas de ASR. As ferramentas submetidas à realização dos testes foram as da Nuance, Google e api.ai, como já tinha sido mencionado anteriormente⁵⁴. O número de amostras de áudio recolhido e usado no teste às ferramentas de ASR foi de 525 amostras referentes a 12 pessoas (5 do sexo masculino e 7 do sexo feminino).

Toolkit de benchmarking

A *toolkit* utilizada para a realização do *benchmarking* foi o HTK⁵⁵, mais especificamente, a ferramenta de análise *HRESULTS*⁵⁶. Esta ferramenta permite-nos, a partir da transcrição de um conjunto de gravações, realizar a comparação com as suas frases originais. O uso desta *toolkit* permite-nos uma forma rápida de avaliação de ferramentas de ASR mesmo com pouco conhecimento dos mecanismos necessários de cálculo e comparação de *strings*.

Para a comparação, é realizado o alinhamento das duas transcrições (frase original e transcrição da gravação) e é, posteriormente, calculado o número de substituições, eliminações e inserções necessárias para a transformação da transcrição da gravação obtida na frase original (ELRA, 2016). Assim, o resultado da comparação entre duas frases é apresentado pela ferramenta no formato ilustrado na Figura 29.

```
===== HTK Results Analysis =====
Date: Tue Jun 21 22:33:47 2016
Ref : E:\Estagio\testesASR\Nuance\sentence_1\sample_1\1_hypothesisMLF.mlf
Rec : E:\Estagio\testesASR\Nuance\sentence_1\sample_1\1_reference.mlf
----- Overall Results -----
SENT: %Correct=0.00 [H=0, S=1, N=1]
WORD: %Corr=78.57, Acc=71.43 [H=11, D=0, S=3, I=1, N=14]
=====
```

Figura 29 - Resultado obtido do HTK do processo comparação de duas transcrições (original e a obtida do ASR)

Da análise da figura, importa destacar da análise do resultado o seguinte:

- O resultado da primeira linha refere-se à eficácia no acerto de transcrições de gravações com as frases originais (número de transcrições exatamente iguais às frases originais).

⁵⁴ Secção 3.2.4

⁵⁵ <http://htk.eng.cam.ac.uk/>

⁵⁶ <http://www.ee.columbia.edu/ln/LabROSA/doc/HTKBook21/node142.html>

- O resultado da segunda linha diz respeito à eficácia de acerto de palavras da transcrição com a frase original, onde são calculadas as métricas referentes ao número de palavras acertadas (H), número de substituições (S), inserções (I) e eliminações (D) a realizar na frase. No entanto, a avaliação da performance é realizada tendo em conta as métricas **%Correct** e **Acc** que resultam da conjugação anteriores (Young, et al., 2009). Elas são expressas pela Equação 1 e Equação 2.

Equação 1 - Percentagem de acerto no reconhecimento correto de palavras

$$\%Correct = \frac{H}{N} * 100\%$$

Equação 2 - Eficácia de reconhecimento de palavras na frase

$$Acc = \frac{H - I}{N} * 100\%$$

Contudo, apesar do resultado apresentado, nenhuma destas métricas é respetiva ao WER que é desejado. A transformação a realizar para se obter o WER é, na realidade, o “inverso” da métrica de *Accuracy*, como se pode ver na Equação 3.

Equação 3 - Cálculo do word error rate (WER) a partir da Accuracy

$$\%WER = 100 - Acc$$

Análise de Resultados

Dos testes realizados pode-se verificar que na sua globalidade a ferramenta da Google tem melhor performance que a da Nuance, como se pode verificar na Figura 30 e Figura 33. Com efeito, da análise da mesma figura é possível verificar mesmo uma maior estabilidade global na média de erro por palavra (WER) da parte da Google comparativamente à Nuance.

Assim, com o intuito de entender o porquê existirem taxas de erro tão elevadas e distintas entre as ferramentas, procedeu-se a uma análise mais detalhada do áudio referente às frases que apresentam maior discrepância no comparativo. O objetivo desta análise é o de descobrir possíveis fatores que podem contribuir para o mau reconhecimento pelas ferramentas de ASR.

Da análise efetuada, verificou-se que ambas as ferramentas apresentam uma média de erro (*avg. WER*) mais baixa, ou seja melhor reconhecimento do conteúdo de gravado, quando se trata de pessoas do sexo do feminino, sendo mesmo os seus valores bastante semelhantes, como se pode ver na Figura 31. Já a respeito do sexo masculino, a Nuance apresenta resultados bastante piores

que a Google que, por seu lado, consegue manter resultados médios de pouco acima dos 50% de erro.

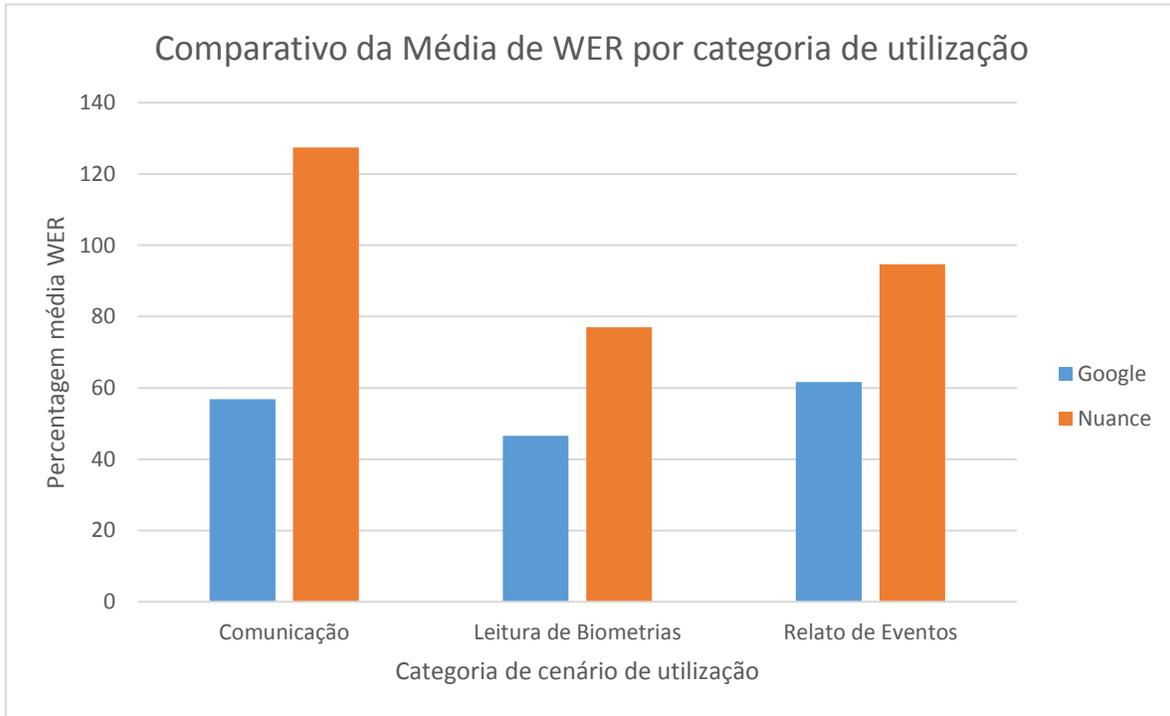


Figura 30 - Comparativo da percentagem média de WER por categoria de cenário de utilização

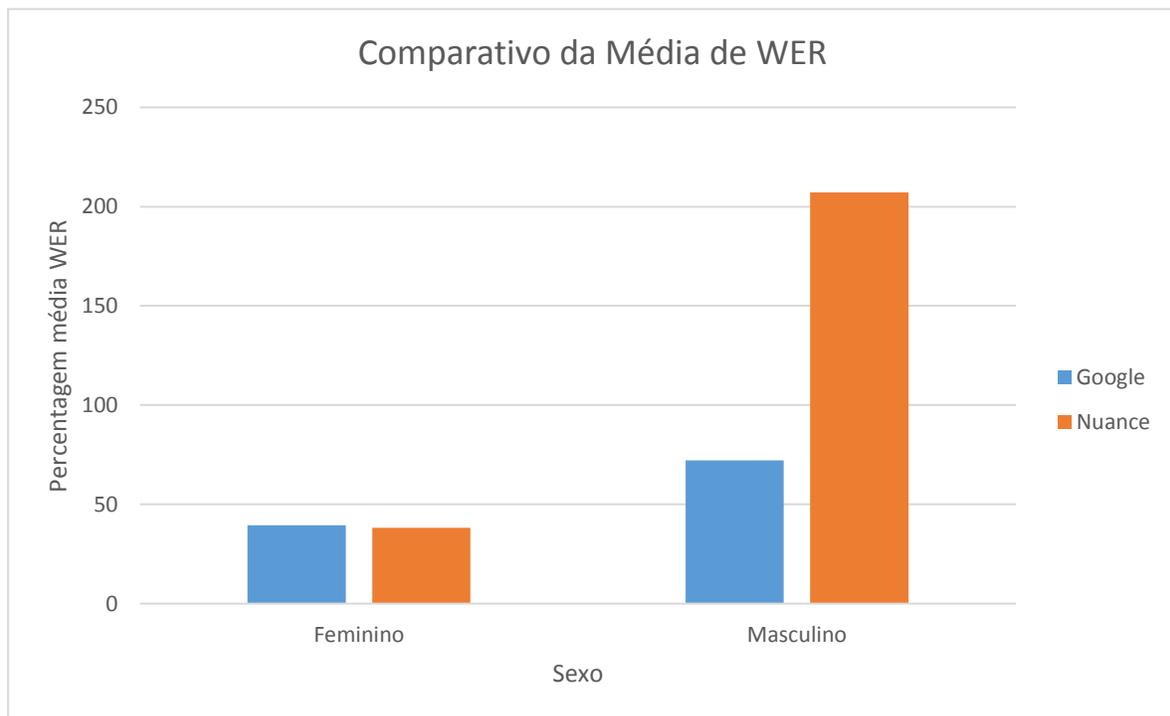


Figura 31 - Comparativo da percentagem média de WER para os dois sexos (Masculino/Feminino)

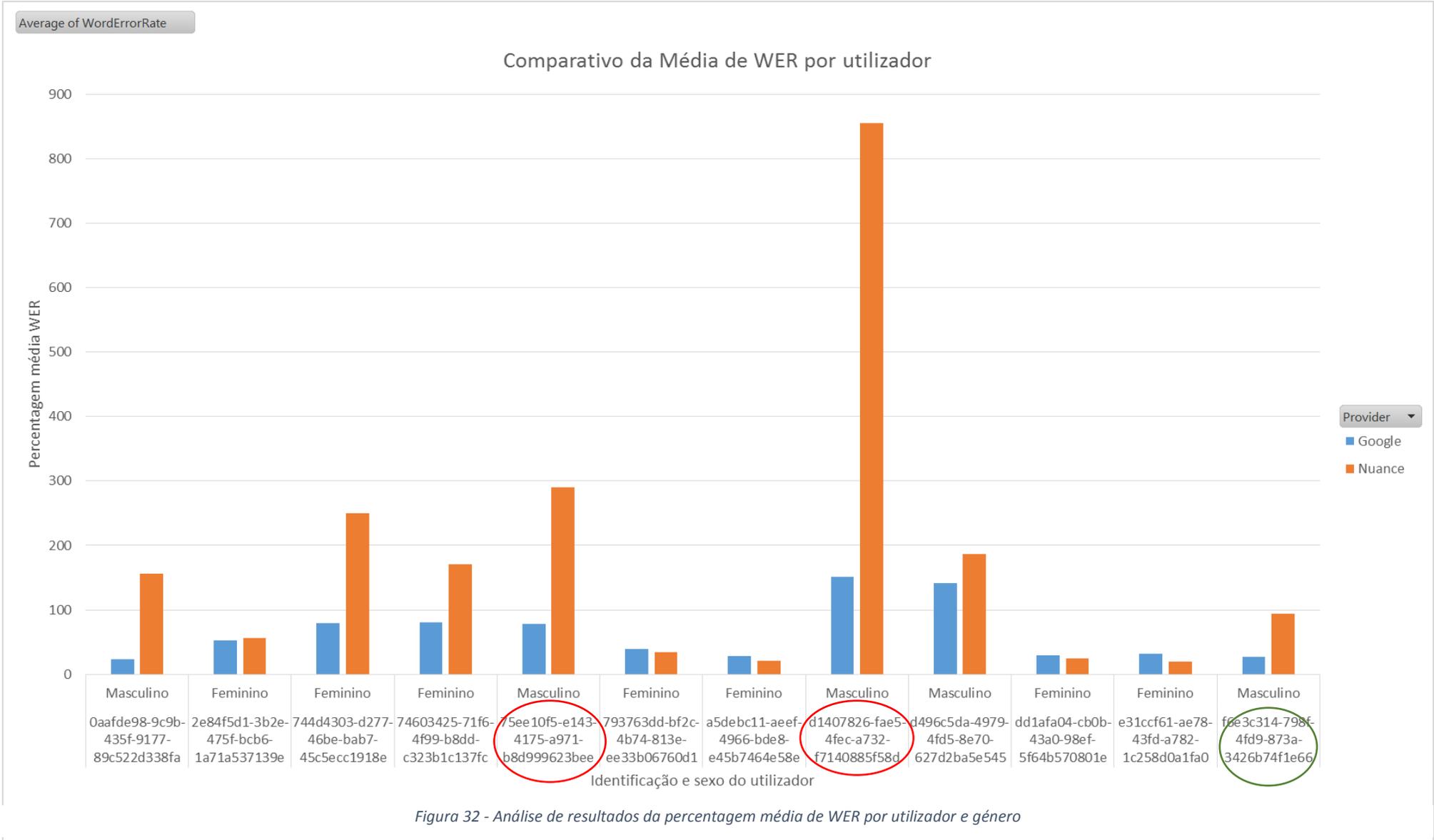
Dado a grande discrepância da percentagem média do WER nos diferentes sexos, procurou-se investigar que características do áudio ou das pessoas podem influenciar os resultados obtidos. Desta forma, foi realizada a audição das gravações recolhidas dos utilizadores que apresentam pior classificação e comparada, posteriormente, com os que melhor destacaram. A análise foi realizada tendo em conta os resultados ilustrados na Figura 32. As características que mais sobressaíram foram as seguintes:

- Timbre da voz (Agudo/Grave)
- Intensidade (Baixa/Normal/Alta)
- Ritmo de dicção

Assim, analisando os dois utilizadores destacados na figura a vermelho verificou-se que ambos eram pessoas com um timbre de voz grave e apesar da clara diferença de intensidade na projeção da voz (primeiro destacado tem menor intensidade). Por outro lado, o utilizador destacado a verde já possuía uma voz intermédia (mais “aguda” que os dois anteriormente mencionados) juntamente com uma projeção maior da voz (mais intensidade). Em termos de ritmo de discurso, à exceção do segundo utilizador destacado a vermelho que teve discurso bastante mais rápido, todos fizeram as gravações de forma natural e pausada.

Para além da análise efetuada a estes utilizadores, verificou-se em algumas gravações analisadas que a presença de ruído ambiente ou digital o que, por sua vez, contribuiu para um pior reconhecimento por parte da Nuance. Por seu lado, a Google no mesmo conjunto de gravações conseguiu lidar com essa presença e obter um melhor resultado.

,



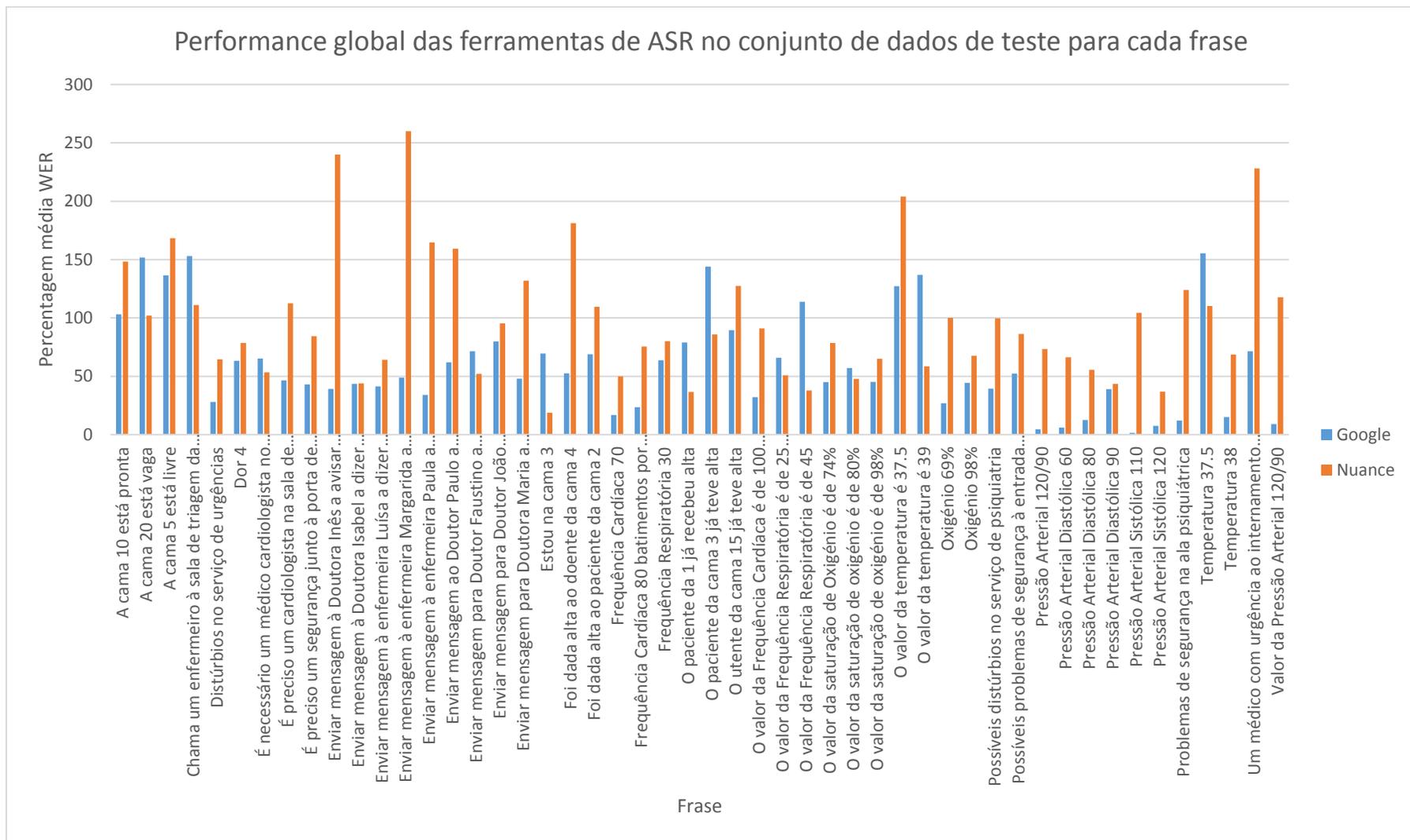


Figura 33 - Avaliação global do WER nos dados de teste para cada uma das ferramentas de ASR (Google e Nuance)

Discussão

Como se pode constatar da análise de resultados dos testes realizados às ferramentas a Google apresenta melhores resultados na maioria dos cenários e utilizadores testados e, na sua maioria, com percentagens médias de erro bastante abaixo que as da Nuance. Além disso, esta análise permitiu identificar certas fraquezas e vantagens que cada uma das ferramentas tem no reconhecimento correto do conteúdo do áudio.

Com efeito, pode-se constatar que a ferramenta da Google consegue garantir uma maior robustez no reconhecimento quando sujeita a uma maior variedade das condições que envolvem o processo de gravação (ex. presença dos diferentes tipos de ruído) que, garantidamente, farão parte de muitos dos ambientes alvo de utilização da solução. Outra virtude também observada foi a maior estabilidade e disponibilidade da ferramenta no suporte às diferentes qualidades da voz (variedade no timbre e intensidade) e intrínsecas de cada indivíduo (ritmo de discurso e sexo).

Por seu lado, a Nuance apresenta-se mais restrita no que diz respeito a estas características. De fato, da análise realizada comporta-se melhor no reconhecimento de vozes do sexo feminino e com certas limitações em lidar com condições adversas no áudio fornecido (presença de diferentes tipos de ruído).

Em suma, do estudo realizado pode-se verificar que, apesar de não serem perfeitos os resultados, estas ferramentas estão a atingir um elevado grau de robustez e que possivelmente, nos próximos 3 a 5 anos, teremos ferramentas quase perfeitas. Todavia, é preciso salvaguardar que, apesar dos resultados obtidos, esta análise foi realizada em ambientes mais controlados sendo necessária a realização de mais testes com mais variedade utilizadores para uma melhor avaliação em cenários reais.

8 Considerações Finais

Para concluir, no presente capítulo é apresentado o ponto de situação atual do trabalho desenvolvido, identificando os principais obstáculos que surgiram durante o período de estágio e qual o trabalho futuro a ser realizado.

8.1 Ponto de situação

Os requisitos funcionais definidos para o âmbito do presente estágio estão todos desenvolvidos no que diz respeito à integração no nível de B2C. Contudo, ao nível da integração B2B os requisitos *Nice* e *Wishful* ficaram por implementar. Isto porque o suporte a estes requisitos neste nível de integração envolve a criação e execução dos processos de negócio respetivos. Por outro lado, ao nível B2C os requisitos são concretizados através dos processos de descoberta de intenções e realização de ações que, por sua vez, alimentam os processos de integração B2B.

Relativamente à solução desenvolvida é de realçar as suas potencialidades no que diz respeito à possível adaptação da mesma a qualquer área ou processos de negócio uma vez que a sua arquitetura

8.2 Principais obstáculos

No período no qual decorreu o estágio houveram 4 principais obstáculos que, devido à sua importância, condicionaram o desenvolvimento de algumas das tarefas inicialmente previstas. Os obstáculos que tiveram mais impacto foram: a in experiência do estagiário nas tecnologias usadas na empresa e nas metodologias formais de teste às ferramentas ASR, a alteração aos requisitos do sistema; múltiplas áreas de estudo e desenvolvimento que envolveram o projeto; e, por fim, a demora na resposta/aprovação de acesso a ferramentas de ASR e NLU da parte algumas empresas.

Esta influência refletiu-se no atraso e, conseqüente, priorização da realização de alguns trabalhos previstos e no definição de um maior período de estudo das tecnologias vigentes na empresa. Por sua vez, isto resultou na falha de análise de todas as tecnologias associadas à implementação de componentes ao nível da integração de processos de negócio (Schmutz, Liebhart, & Welkenbach,

2010) – através do uso de tecnologias como BizTalk Server⁵⁷⁵⁸, *Enterprise Service Bus* (ESB) – e a experimentação de outros serviços de processamento semântico, como por exemplo o *Mix.NLU* e o *360 Development Platform* da Nuance.

De realçar que, dada a natureza do presente projeto, envolveu o trabalho em diferentes áreas de desenvolvimento que, num cenário de real, nunca seria realizado por uma mesma pessoa num mesmo projeto dada a complexidade que caracteriza cada uma das áreas. Para além disso, o fato de não haver nenhuma pessoa especializada ou sequer com o mínimo de conhecimento relativo ao desenvolvimento nestas áreas na empresa, como são os casos das áreas de processamento de áudio, NLU e testes (tanto às ferramentas de ASR como de integração), fez com que muito mais tempo tivesse que ser despendido para a realização de cada uma das tarefas envolvidas. Isto levou a que a análise realizada dos testes realizados na secção 7.4 não fosse profunda o suficiente na avaliação dos fatores que podem ou não contribuir para o correto reconhecimento do discurso do utilizador.

8.3 Trabalho futuro

Como se pode constatar da leitura do documento, apesar de que foi conseguido, existem ainda bastantes pontos que se podem melhorar no que diz respeito no teste e análise das ferramentas de ASR e NLU existentes.

De fato, no que diz respeito às ferramentas de ASR, dadas a recente disponibilização da API da Google e a descoberta de outras ferramentas não encontradas no processo de estado da arte que suportam PT-PT pode-se impor uma nova ronda de testes tendo em conta estas alternativas com o objetivo de realizar uma melhor escolha. São exemplos destas ferramentas a plataforma VoxSigma^{59,60} da Vocapia Research e a solução SayIt da nVoq Inc. que, por sua vez, tanto fornece soluções de ASR como também possui ferramentas de automatização de fluxos de trabalho e de ditado, entre muitas outras.

Para além disso, de forma a obter resultados de performance mais realistas e contextualizados, impõe-se a realização de testes nos diferentes ambientes alvo em que se pretenda no futuro usar a ferramenta de ASR como mecanismo principal de interação dos utilizadores. Isto porque a avaliação

⁵⁷ [https://msdn.microsoft.com/en-us/library/ee267544\(v=bts.10\).aspx](https://msdn.microsoft.com/en-us/library/ee267544(v=bts.10).aspx)

⁵⁸ [https://msdn.microsoft.com/en-us/library/ee251575\(v=bts.10\).aspx](https://msdn.microsoft.com/en-us/library/ee251575(v=bts.10).aspx)

⁵⁹ <http://www.vocapia.com/speech-to-text-api.html>

⁶⁰ <http://www.vocapia.com/speech-to-text-services.html>

da performance da ferramenta nestes ambientes estará mais próxima das condições reais de uso de uma aplicação.

Por outro lado, importa também destacar a possibilidade de se realizarem testes formais às ferramentas de NLU com o intuito de avaliar outras alternativas à api.ai ainda que, na minha opinião, esta ser claramente muito superior às restantes ferramentas estudadas e fornecer suporte bastante personalizado durante o desenvolvimento e, não esquecendo, ser das únicas que já suporta NLU na língua portuguesa (além das mencionadas na secção 3.3).

Para concluir, seria interessante também realçar a possibilidade de realização de testes de usabilidade às APIs desenvolvidas (em inglês *API Usability Testing*)⁶¹ para os diferentes componentes como forma de avaliar a experiência de utilização e perceção dos programadores que querem integrar os componentes desenvolvidos noutros sistemas e/ou aplicações.

⁶¹ <http://18f.github.io/API-Usability-Testing/>

Bibliografia

- (s.d.). Obtido de <https://kishore1021.files.wordpress.com/2010/07/mefppt2010.jpg>
- (28 de Maio de 2016). Obtido de Wikipedia: <https://pt.wikipedia.org/wiki/Stakeholder>
- (8 de Junho de 2016). Obtido de AT&T Developer Program:
<http://developer.att.com/apis/speech/docs/v3>
- (12 de Junho de 2016). Obtido de Himms:
<http://exhibitionfloor.himss.org/himss2016/public/eBooth.aspx?IndexInList=576&FromPage=Exhibitors.aspx&ParentBoothID=&ListByBooth=true&BoothID=167484&Nav=False>
- (12 de Junho de 2016). Obtido de About WordNet: <https://wordnet.princeton.edu/>
- Ammann, P., & Offutt, J. (2008). *Introduction to Software Testing*. Cambridge University Press.
- Bass, L., Clements, P., & Kazman, R. (2003). *Software Architecture in Practice*. Addison-Wesley.
- Broker*. (11 de Junho de 2016). Obtido de Msdn Microsoft: <https://msdn.microsoft.com/en-us/library/ff648096.aspx>
- Brokered Authentication: Security Token Service (STS)*. (6 de Junho de 2016). Obtido de Msdn Microsoft: <https://msdn.microsoft.com/en-us/library/ff650503.aspx>
- Clements, P., Bachman, F., Bass, L., Garlan, D., Ivers, J., Little, R., . . . Stafford, J. (2003). *Documenting Software Architectures: Views and Beyond*. Addison-Wesley.
- ELRA. (21 de Junho de 2016). Obtido de ELRA European Language Resources Association:
<http://www.hlt-evaluation.org/spip.php?article147>
- Employee Data Privacy Policy: European Union*. (8 de Junho de 2016). Obtido de Nuance:
<http://www.nuance.com/company/company-overview/company-policies/safe-harbor/index.htm>
- Fellbaum, C. (1998). *WordNet*. Wiley Online Library.
- Gaida, C., Lange, P., Petrick, R., Proba, P., Malatawy, A., & Suenderman-Oeft, D. (2014). Comparing Open-Source Speech Recognition Toolkits. *Oasis*.
- Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley.
- Hohpe, G., & Woolf, B. (2004). *Enterprise Entegration*. Addison-Wesley.
- Hunt, J. (2006). *Agile Software Construction*. Springer Science and Business Media.
- Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall.
- Lobo, J. P. (Junho de 2013). Personal Medical Advisor.

- Martin, R. C. (2002). *Agile Software Development Principles, Patterns, and Practices*. Prentice Hall.
- Martin, R. C. (22 de Maio de 2016). Obtido de Clean Coder Blog:
<http://blog.cleancoder.com/uncle-bob/2016/01/04/ALittleArchitecture.html>
- Morbini, F., Audhkhasi, K., Sagae, K., Artstein, R., Can, D., Georgiou, P., . . . Traum, D. (2013). Which ASR should I choose for my dialogue system? *Proceedings of the SIGDIAL 2013 Conference* (p. 10). Metz, France: Association for Computation Linguistics.
- Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys (CSUR)*, p. 10.
- Roshen, W. (2009). *SOA-Based Enterprise Integrations: A Step-by-Step Guide to Services-based Application*. McGraw-Hill, Inc.
- Schmutz, G., Liebhart, D., & Welkenbach, P. (2010). *Service-Oriented Architecture: An Integration Blueprint*. Packt Publishing.
- The Repository Pattern*. (1 de Junho de 2016). Obtido de msdn: <https://msdn.microsoft.com/en-us/library/ff649690.aspx>
- Wieggers, K., & Beatty, J. (2013). *Software Requirements*. Microsoft Press.
- Wolber, A. (12 de Junho de 2016). *How to Configure Google Apps for HIPAA Compliance*. Obtido de datto: <http://www.datto.com/blog/configure-google-apps-hipaa-compliance>
- Young, S., Evermann, G., Gales, M., Hain, T., Liu, X. (., Odell, . . . Woodland, P. (2009). *The HTK Book*. Cambridge University Engineering Department.