Identificação de características de sistemas legados a partir da análise de conteúdo da literatura

Daniele Martins, Alex Chervenski, Andréa S. Bordin

Curso de Bacharelado em Engenharia de Software – Universidade Federal do Pampa (UNIPAMPA) – Campus Alegrete
97546-550 – Alegrete– RS– Brasil

{danimarttiins,alex.chervenski}@gmail.com,andreabordin@unipampa.edu.br

Abstract. *Legacy systems tend to consume a lot of resources and so, requires to be well managed. Deciding whether a system is becoming legacy or not is a decision that involves identifying the characteristics of such a system. This paper aims to present the mapping of features found in the literature's definitions of legacy systems. For this, a systematic process of search, extraction and analysis of the definitions of legacy systems found between the years 1995, 2000, 2005, 2010 and 2015 was performed. The results analysis reveal that in the studied period, the characteristics with the highest frequency of occurrences were, to be vital to the organization, to be a obsolete technology and to have no documentation.*

Resumo. *Sistemas legados tendem a consumir muitos recursos e por isso devem ser bem gerenciados. Decidir se um sistema é ou está se tornando legado é uma decisão que envolve a identificação das características de tal sistema. Este artigo tem como objetivo apresentar o mapeamento de características encontradas nas definições de sistemas legados utilizadas na literatura. Para isso foi realizado processo sistemático de busca, extração e análise de conteúdo das definições de sistemas legados encontradas nos anos de 1995, 2000, 2005, 2010 e 2015. A análise dos resultados revela que as características relacionadas ao uso de tecnologia obsoleta, ser vital para a organização e não possuir documentação são as que possuem maior frequência de ocorrências no período pesquisado.*

1. Introdução

[Ning et al. 1994] descreve sistemas legados como sistemas que inibem o crescimento de uma organização e sua capacidade de mudança, [Bennett 1995] diz que são sistemas que realizam tarefas úteis para a organização, mas foram desenvolvidos com tecnologias atualmente consideradas obsoletas. Outra definição diz que sistemas legados "são sistemas com valor crítico para o negócio das empresas desenvolvidos há cinco ou mais anos" [Umar 1997]. Por outro lado, [Juric et al. 2000] definem que "é qualquer sistema que, independentemente da idade ou arquitetura, ainda é útil e está em

uso".

As quatro definições mostram diferentes características e evidenciam que não há um consenso sobre o que é um sistema legado. De acordo com [O'Byrne e Wu 2000] não saber ao certo quais são essas características torna difícil o processo de identificar se um sistema é ou está se tornando realmente "legado". Ainda, segundo esses autores, é difícil determinar com precisão se um sistema é ou não é legado, logo a pergunta mais adequada poderia ser "Em qual o grau de sistema legado um determinado sistema se encontra? ".

A identificação desse tipo de sistema é importante para uma posterior avaliação e tomada de decisão sobre o que fazer com o sistema. Este é um grande desafio para as organizações, visto que tais sistemas costumam consumir muitos recursos para mantê-los. [Ramsom et al. 1998] e [Sommerville 2011] apontam que possíveis ações, resultantes de um processo de avaliação de sistemas legados, incluem continuar com a manutenção, reengenhar o sistema, substituir ou aposentar o sistema.

Este artigo objetiva apresentar o mapeamento das características encontradas com mais frequência nas definições de sistemas legados utilizadas na literatura. Os resultados desse mapeamento apontam as características mais utilizadas pelos autores, as quais por sua vez, podem ser utilizadas como referência para a identificação de sistemas que estão se tornando ou já são sistemas legados.

2. Metodologia

O processo sistemático de busca de definições de sistemas legados ocorreu entre maio a setembro de 2016 e foi conduzido com base nas etapas de planejamento e execução de revisão sistemática propostas por [Kitchenham 2004]. Na fase de planejamento foi identificada a inexistência de pesquisas relacionadas ao mapeamento de definições de sistemas legados e foi elaborado um protocolo no qual foram definidas a questão de pesquisa, a estratégia de busca e os critérios de inclusão e exclusão dos estudos.

A questão de pesquisa definida foi: "Quais as definições de sistema legados utilizadas na literatura?". A estratégia de busca e os critérios são apresentadas na Tabela 1:

Tabela 1. Estratégia de busca e Critérios de Inclusão e Exclusão

Fonte de busca	Google Acadêmico
Idioma de estudo	Inglês
String de busca	("legacy system is") OR ("legacy software is") OR

	("legacy systems are") OR ("legacy software are")
Critérios de Inclusão (CI)	(a) Conter uma definição de sistemas legados no abstract ou no corpo do artigo; (b) Ter sido publicado no ano de 1995 ou 2000 ou 2005 ou 2010 ou 2015.
Critérios de Exclusão (CE)	Estudos não disponíveis em texto completo.

A fonte de busca foi escolhida em função da abrangência de indexação de estudos e porque outras bases de dados retornaram baixo número de estudos nas buscas preliminares à essa revisão. O intervalo de cinco anos foi determinado em função da análise da variação do conceito de sistemas legados ao longo do tempo

Foram encontrados no período de 1995 até 2015 um total de 7.600 artigos, os quais foram lidos e selecionados através dos critérios de inclusão e exclusão. O número de artigos resultantes após a aplicação do CI (b) é exibido na Tabela 2.

Tabela 2. Número de artigos antes e depois dos CI e CE

Ano	Antes dos CI e CE	Depois dos CE e CI
1995	88	6
2000	268	17
2005	360	24
2010	445	22
2015	388	31
Total	**1549**	**100**

Um total de 125 definições de sistemas legados foram extraídas manualmente e armazenadas a para posterior análise individual de cada definição. Foram considerados os seguintes tópicos: (a) as definições de sistemas legados; (b) a autoria do artigo; (c) o endereço do repositório; (d) o ano de publicação. A soma das definições supera a quantidade de artigos selecionados porque alguns estudos apresentaram mais de uma definição distinta para sistemas legados.

Utilizou-se a técnica de análise de conteúdo para o tratamento das definições

extraídas dos estudos. Para [Moraes 1999] a análise de conteúdo tem como objetivo "descrever e interpretar o conteúdo de toda classe de documentos e textos obtidos através de uma pesquisa.". O processo de análise de conteúdo possui cinco etapas: preparação das informações; unitarização do conteúdo; categorização das unidades; descrição das categorias e interpretação dos resultados.

Na etapa de unitarização do conteúdo cada definição foi analisada detalhadamente em dois momentos, individualmente e em grupo, com o objetivo de extrair a unidade de análise, neste artigo denominada de característica. Como resultado desse processo, as principais características de cada uma das definições e respectivas frequências foram extraídas e armazenadas. Ao final, das 125 definições de sistemas legados analisadas, geraram-se 62 unidades de análise ou características. Optou-se por utilizar somente as definições que apresentavam frequência maior ou igual a 2, de forma que neste trabalho apresenta-se as 31 características principais.

Logo após iniciou-se a fase de categorização, onde as unidades com características semelhantes foram agrupadas em categorias aqui denominadas de Aspectos. Por exemplo, a característica "Desenvolvido há muito tempo" foi agrupada com a característica "Desenvolvido há mais de 20 anos", formando o Aspecto de Tempo. O resultado deste agrupamento gerou um total de sete categorias identificadas como: "Aspectos de Estrutura de Software", "Aspectos de Manutenção", "Aspectos de Recursos Humanos", "Aspectos de Tecnologia Obsoleta", "Aspectos de Tempo", "Aspectos Financeiros" e "Aspectos Organizacionais". A seguir, na Tabela 3 os seus significados são brevemente explicados.

Tabela 3. Descrição dos Aspectos

Aspecto	Descrição
Estrutura de Software	Engloba características como alta complexidade, ausência de documentação e considerável número de linhas de código, o que torna a arquitetura inflexível a modificações e desatualizada para competir com equivalentes modernos.
Manutenção	Considera características relacionadas à necessidade de extensivas manutenções ao longo de seu ciclo de vida, à lentidão e custo das manutenções.
Recursos Humanos	Refere-se a características relacionadas a inexistência de mão-de-obra especializada ou indisponibilidade dos desenvolvedores originais dos sistemas, o que inviabiliza modificações no software.

Tecnologia obsoleta	Características de sistemas desenvolvidos para hardware antigo e linguagens de programação que não são mais utilizadas. O termo tecnologia inclui hardware, software e técnicas utilizadas para a construção de tais sistemas.
Tempo	Características relacionadas ao período do tempo em que o sistema foi desenvolvido, num passado remoto ou próximo.
Financeiro	Características relacionadas ao valor econômico investido pela organização ao longo dos anos.
Organizacionais	Refere-se a características que indicam a importância do sistema para a organização.

3. Resultados e Discussão

A Figura 1 exibe as características mais presentes nas definições de sistemas legados classificadas por Aspectos e com suas respectivas frequências absolutas, cujo total é 324.

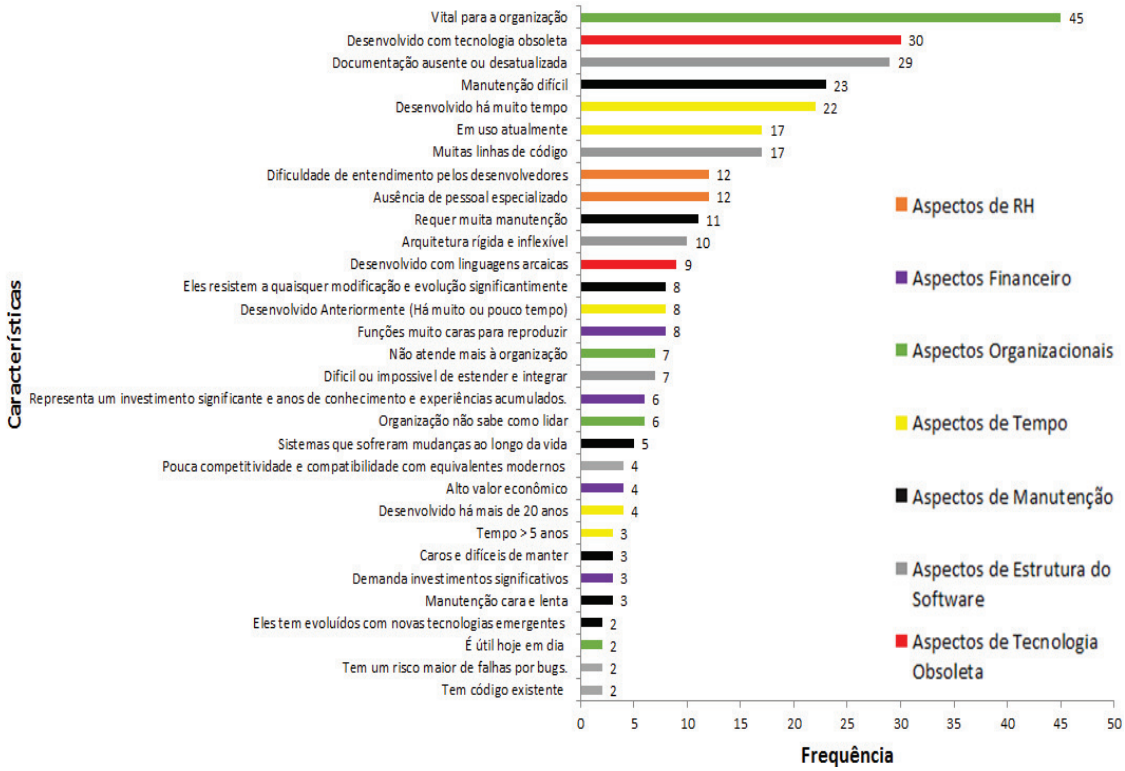
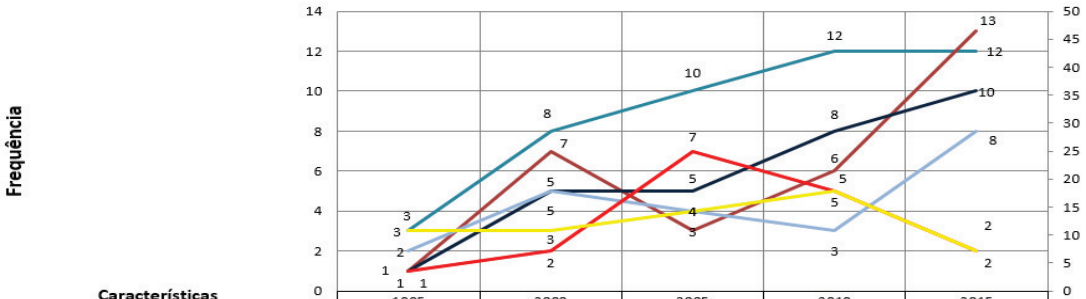


Figura 1. Frequência Total de Características de Sistemas Legados

Observa-se que a característica com maior frequência relativa é "Vital para organização" com 13%, uma característica não tecnológica que evidencia o papel fundamental desse tipo de sistema para as organizações. Na sequência, aparecem as características "Desenvolvido com tecnologia obsoleta" com 9%, "Documentação ausente ou desatualizada" com 8,9%, "Manutenção difícil" com 7% e "Desenvolvido há muito tempo" com 6,7%, às quais já são amplamente conhecidas e imediatamente relacionadas à ideia de sistema legado.

As características presentes nas definições de sistemas legados sofreram variações ao longo do período pesquisado. A evolução da frequência das características encontradas simultaneamente em todos os anos estudados é visualizada na Figura 2. A partir de uma visão global observa-se essas características aparecem em todos os anos do período pesquisado já pode ser considerado um indicador significativo da importância das mesmas. Destaca-se ainda que a frequência de todas as características aumenta a cada ano do período pois também aumenta o número de artigos publicados com essa temática (Tabela 2).



Características	1995	2000	2005	2010	2015
Vital para a organização	3	8	10	12	12
Desenvolvido com tecnologias obsoletas	1	7	3	6	13
Documentação ausente ou desatualizada	1	5	5	8	10
Desenvolvido há muito tempo	2	5	4	3	8
Em uso atualmente	1	2	7	5	2
Muitas linhas de código	3	3	4	5	2
TOTAL	11	30	33	39	47

Figura 2. Evolução de características no intervalo dos anos estudados.

Em uma perspectiva local, a característica 'Vital para a organização' foi a mais frequente e manteve uma evolução crescente nos quatro primeiros anos pesquisados, o que demonstra uma conscientização crescente da importância desse tipo nas organizações. Já a característica 'Desenvolvido com tecnologia obsoleta' apareceu de forma inconstante nas definições ao longo dos anos, sendo citada apenas uma vez em 1995 e treze vezes em 2015, sendo, portanto, a característica mais citada no último período pesquisado. Esse aumento significativo no último ano deve estar relacionado à própria evolução das tecnologias de informação, que faz com que novos recursos de hardware e software se tornem disponíveis, tornando as tecnologias anteriores

rapidamente obsoletas. Outra característica com evolução considerável foi 'Documentação ausente ou desatualizada' que em 1995 apareceu em somente uma definição, passando para dez vezes em 2015. Aqui o motivo pode estar relacionado ao aumento na demanda de manutenção de sistemas legados, pois é senso comum que muitos sistemas, principalmente os desenvolvidos no passado, carecem de documentação.

Como mencionado na metodologia, a fase de categorização agrupou as características em categorias denominadas Aspectos. Sendo assim, visualiza-se na Figura 3 os Aspectos com suas respectivas frequências em ordem decrescente.

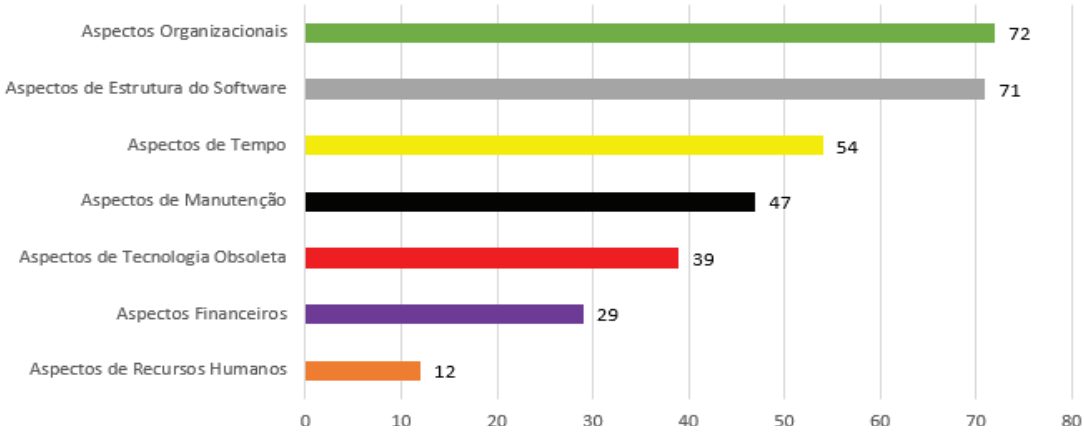


Figura 3. Agrupamento de características em Aspectos

Ao analisar a Figura 3, nota-se a elevada frequência dos Aspectos Organizacionais e dos Aspectos de Estrutura do Software. Essa situação parece refletir que tanto aspectos técnicos quanto aspectos sociais e organizacionais estão envolvidos no escopo de sistemas legados.

4. Ameaças à validade

Neste estudo é considerada uma ameaça à validade a interpretação das características encontradas nas definições de sistemas legados, que pode conduzir a equívocos na contabilização das frequências das características e à atribuição equivocada da característica a um aspecto.

5. Conclusão

O mapeamento de um conjunto consolidado de características de sistemas legados é o propósito principal desta pesquisa. O processo de busca sistemática resultou em 100 artigos, a partir dos quais e, através da análise de conteúdo, foram identificadas 31 características diferentes, que por fim foram agrupadas em 7 aspectos.

A identificação revelou um conjunto considerável de características, com

destaque, em termos de frequência absoluta, para a característica 'Vital para a organização'. Ao analisar as demais características com frequência maior que 20 é possível propor que um sistema legado seja 'Um sistema vital para organização, desenvolvido com tecnologia obsoleta, com documentação ausente ou desatualizada, de difícil manutenção e desenvolvido há muito tempo'.

Conclui-se que a utilização da abordagem de busca sistemática em conjunto com a técnica de análise de conteúdo mostrou-se relevante para a obtenção do conjunto final de características e categorias que representam o conceito de sistema legado do ponto de vista acadêmico ao longo de 20 anos e podem contribuir para facilitar a identificação de tais sistemas nas organizações e, com isso, permitir que a melhor decisão em relação a esses sistemas seja tomada.

Como trabalhos futuros, pretende-se realizar a busca sistemática em todos os anos, com o propósito de tornar os resultados mais robustos e, a partir do conjunto final de características e aspectos, propor a construção de uma taxonomia de características de sistemas legados com o objetivo de organizar o conhecimento acerca deste tópico.

Referências

Bennett, K. (1995). Legacy systems: Coping with success. IEEE Software, 12(1):19–23.

Juric, M. B., Rozman, I., and Hericko, M. (2000). Performance comparison of corba and RMI. Information and Software Technology, 42(13):915–933.

Kitchenham, B. (2004). Procedures for performing systematic reviews. Technical report, Keele University, Keele, UK, Software Engineering Group. Moraes, R. (1999). Análise de conteúdo. Revista Educação, 22(37):7–32.

Ning, J., Engberts, A. P., and Kozaczynski, W. V. (1994). Automated support for legacy code understanding. Communications of the ACM, 37(5):50–57.

O'Byrne, P. and Wu, B. (2000). Lace frameworks and technique - identifying the legacy status of an information system from the perspectives of its causes and effects. In Proceedings International Symposium on Principles of Software Evolution.

Ramsom, J., Sommerville, I., and Warren, I. (1998). A method for assessing legacy systems for evolution. Proceedings of the Second Euromicro Conference on Software Maintenance and Reengineering, 42(13):915–933.

Sommerville, I. (2011). Engenharia de Software. Pearson Brasil, 9th edition.

Umar, A. (1997). Application (re)engineering: Building Web-based Applications and Dealing with Legacies. Prentice Hall, 1th edition.