

Uma Webibliometria Sobre a Relevância de Arquiteturas Computacionais

Diogo Valadares Reis dos Santos*, David Vasconcelos Corrêa da Silva

2025

Resumo

Este artigo explora a aplicação educacional de arquiteturas computacionais por meio de uma análise bibliométrica para avaliar a relevância das arquiteturas RISC no contexto mais amplo dos sistemas de computação. Utilizando ferramentas como o Google Books Ngram Viewer e bases de dados acadêmicas como o Google Scholar e os periódicos da CAPES, o estudo identifica tendências e padrões na evolução e adoção de arquiteturas computacionais. As principais descobertas incluem a proeminência cronológica dos designs baseados em RISC e sua influência em arquiteturas modernas, como o RISC-V. A pesquisa combina dados de uso histórico com uma avaliação das tendências de publicações acadêmicas, oferecendo uma visão sobre a importância duradoura dessas arquiteturas para avanços educacionais e tecnológicos.

Palavras-chaves: Arquiteturas computacionais, RISC, análise bibliométrica, Ngram Viewer, relevância acadêmica

Introdução

No desenvolvimento do trabalho *DRISC: Uma Implementação Didática da Arquitetura RISC I*, foi necessário realizar uma pesquisa sobre arquiteturas relevantes ao longo dos anos. O objetivo era tanto obter uma perspectiva sobre a importância da arquitetura RISC I em comparação com outras, quanto identificar quais recursos são atualmente considerados relevantes em arquiteturas computacionais.

Este artigo detalha o processo dessa pesquisa, que inclui uma análise bibliométrica utilizando o recurso de análise de *n-grams* (MICHEL et al., 2011) da ferramenta *Google Books Ngram Viewer* (GBNV), além de pesquisas nas bases acadêmicas *Google Scholar* e Periódicos CAPES. Este processo baseia-se no trabalho de Costa (2010), que emprega uma estratégia de coleta e análise de dados bibliográficos disponíveis na web, a fim de obter uma seleção de um núcleo inicial de artigos para uma pesquisa.

*diogo-valadares@hotmail.com

No trabalho de Costa (2010), é descrito um processo para identificar periódicos com maior número de resultados e autores com mais publicações. No entanto, o objetivo desta pesquisa é diferente. A bibliometria está sendo realizada de duas formas distintas. A primeira, como mencionado anteriormente, utiliza uma análise de *n-grams* para obter uma análise cronológica da utilização dos nomes de certas arquiteturas computacionais ao longo dos anos. Essa abordagem permite identificar quais arquiteturas são mais relevantes atualmente e fornece uma lista de arquiteturas para comparação, visando encontrar recursos comuns entre elas.

A segunda abordagem de bibliometria executada é mais semelhante ao trabalho original de Costa, mas com algumas diferenças. Foram realizadas pesquisas no *Google Scholar* e Periódicos CAPES com o nome de cada arquitetura para estimar quantos trabalhos acadêmicos citam essas arquiteturas. Diferente do trabalho de Costa, esta bibliometria não busca valores exatos, mas sim estimativas que ajudem a obter uma ideia geral do que é necessário pesquisar. Essa abordagem permite uma visão mais ampla e flexível das tendências e relevâncias das arquiteturas computacionais ao longo do tempo.

1 Pesquisa Inicial

1.1 Definição da amostra

A amostra de pesquisa inclui tanto os livros em inglês utilizados na base utilizada pelo GBNV, identificado como googlebooks-eng-20200217, quanto aos trabalhos acadêmicos indexados ao Google Scholar e Periódicos CAPES. As pesquisas se limitaram a arquiteturas que foram inventadas a partir dos anos 70 e tiveram um foco maior sobre arquiteturas RISC, mas sem a exclusão de arquiteturas CISC.

1.2 Escolha de arquiteturas

Para escolher quais arquiteturas foram analisadas, foram realizadas pesquisas gerais na internet para a obtenção das mais utilizadas durante os anos. Uma das fontes utilizadas foi o Top 500, acessível em top500.org/statistics/overtime/, que mostra quais foram as arquiteturas utilizadas nos 500 supercomputadores mais potentes dos últimos 30 anos. Essa lista forneceu a confirmação da utilização de diversas arquiteturas, das quais podemos citar PA-RISC, PowerPC, SPARC, x86, x86-64 e MIPS.

Outras arquiteturas foram também consideradas a partir de inspirações para o RISC-V, a última iteração dos processadores Berkeley RISC. Estas podem ser vistas no trabalho de (PATTERSON; CHEN, 2016), onde é feita uma comparação das instruções de diversas arquiteturas. Dentre estas podemos citar os RISCs anteriores, O Intel i960 (pré x86), ARM, DLX, MIPS, PA-RISC, PowerPC e SPARC. Note que muitos dos nomes informados não se referem a uma arquitetura específica, e sim a famílias de arquiteturas, o que será relevante para os passos seguintes.

2 Análise com n-grams

Segundo Michel et al. (2011), um 1-gram é uma sequência de caracteres sem interrupção por um espaço; isso inclui palavras ('banana', 'SCUBA'), mas também números ('3.14159') e erros de digitação ('excesss'). Um n-gram é uma sequência de 1-grams, como as frases 'mercado de ações' (um 2-gram) e 'os Estados Unidos da América' (um 5-gram).

A ferramenta *Google Books Ngram Viewer* permite a visualização da frequência de uso de *n-grams* em uma base de dados composta por 8 milhões de livros digitalizados, representando aproximadamente 6% de todos os livros já publicados (LIN et al., 2012). A frequência de uso é calculada dividindo-se o número de ocorrências do *n-gram* em um determinado ano pelo total de palavras no corpus naquele ano (MICHEL et al., 2011). O GBNV também possibilita a visualização da frequência ao longo do tempo em um gráfico, permitindo analisar qual foi o período mais relevante de uma arquitetura.

2.1 Limitações

O uso dessa ferramenta pode fornecer informações importantes sobre a utilização desses *n-grams* na literatura, porém apresenta algumas limitações. Certos termos podem ser usados em mais de um contexto, o que dificulta a análise de nomes genéricos ou números. Por exemplo, uma das arquiteturas pesquisadas chama-se SPUR, que pode referir-se tanto a um acessório de sapato em inglês quanto a uma sigla para diversas coisas, como ilustrado na Figura 1.

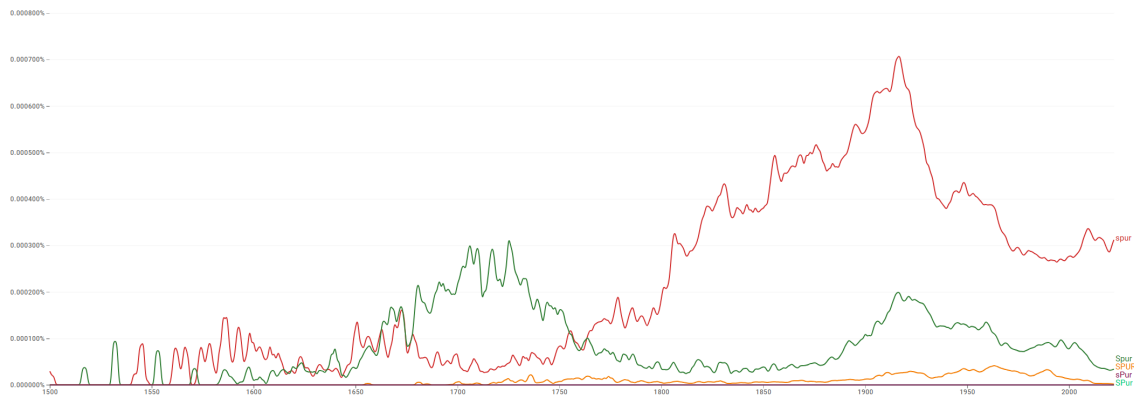


Figura 1 – SPUR no *n-gram viewer* entre 1500 e 2020

Há maneiras de obter dados nessas situações, utilizando termos alternativos que retornem resultados relevantes ao que se busca; adicionando *n-grams* ao objeto de análise para proporcionar contexto, como "SPUR architecture", ou empregando o recurso de subtração para excluir resultados que desviem os valores, como "SPUR - SPUR Association". Embora essas soluções funcionem em muitos casos, elas não produzem resultados tão precisos quanto *n-grams* únicos, o que pode gerar distorções nos resultados.

Outra limitação dessa ferramenta é que ela se restringe ao conteúdo disponível em livros, excluindo outros tipos de fontes que podem representar melhor a relevância de certos *n-grams*, como artigos ou páginas da web. A frequência de *n-grams* também pode não ser uma métrica muito precisa dependendo do objetivo da pesquisa, pois um livro pode usar o mesmo *n-gram* várias vezes.

Um grande problema é que a frequência de uso de um *n-gram* pode ser inflada se uma ou mais obras específicas mencionam o termo muitas vezes. Como a ferramenta calcula a frequência com base no número total de ocorrências em relação ao número total de palavras no corpus (MICHEL et al., 2011), um único livro que utiliza um termo repetidamente pode aumentar desproporcionalmente a frequência de uso, dando uma

impressão errônea da popularidade do termo. Portanto, mesmo que a frequência de uso de um termo seja maior do que a de outro, isso não indica diretamente seu nível de popularidade em comparação com outros termos.

Apesar destas desvantagens, este método é simples de ser executado, podendo fornecer uma rápida estimativa sobre a relevância do que está sendo pesquisado. Além disso, há uma possibilidade de poder pesquisar diversos *n-grams* ao mesmo tempo permite uma rápida comparação entre os períodos em que cada *n-gram* foi mais usado.

3 Análise em Bases Acadêmicas

Para uma comparação direta da popularidade entre as arquiteturas de computadores, um passo adicional pode ser realizado utilizando ferramentas de busca de trabalhos acadêmicos. Este passo adicional oferece uma comparação da quantidade de trabalhos relacionados, em vez de apenas comparar a frequência de uso de termos específicos.

Esta análise pode ser combinada com a realizada no visualizador de Ngrams, proporcionando um entendimento mais amplo da relevância das arquiteturas. Enquanto o Ngram Viewer permite uma visualização da relevância ao longo do tempo, uma análise em uma base de dados oferece uma visão geral da relevância, pois os resultados são baseados em trabalhos publicados, não apenas no uso de termos. Isso possibilita uma comparação direta da relevância entre diferentes arquiteturas.

Para realizar esta análise, foram utilizadas duas bases de dados: o [Google Scholar](#) e o [Periódicos Capes](#). Cada uma dessas bases possui mecanismos de busca e artigos indexados diferentes, o que permite verificar tanto se as técnicas de pesquisa se aplicam para diferentes bases quanto realizar uma comparação para validação dos resultados. Adicionalmente, o Google Scholar permite pesquisas por patentes, aumentando o tipo de documentos incluídos na pesquisa.

3.1 Vantagens e Desvantagens

As pesquisas realizadas dentro das bases acadêmicas apresentam uma métrica de relevância mais precisa, isso por que ao invés de comparar a quantidade de vezes que um termo foi utilizado, estamos comparando a quantidade de trabalhos que apresentam este termo, eliminando problemas em situações onde poucos trabalhos citam o mesmo termo muitas vezes.

A filtragem dos termos também se torna bem mais simples, isso porque podemos combinar os nomes das arquiteturas com outros termos específicos removendo uma quantidade pequena de trabalhos relacionados. Isso se deve ao fato de que a combinação de termos pode diminuir a frequência de uso no geral, o que é ruim para uma análise nos Ngrams, porém dificilmente removerá trabalhos relevantes que devem citar estas combinações ao menos uma vez.

Embora as pesquisas em bases acadêmicas tenham suas vantagens, elas geralmente não oferecem uma forma fácil de análise temporal, limitando-se a filtros anuais. Isso é bastante restritivo em comparação com os Ngrams, que apresentam gráficos de fácil interpretação visual.

Além disso, algumas bases de dados podem não representar com precisão a totalidade dos trabalhos publicados. Por essa razão, é recomendada a utilização de mais de

uma base para pesquisas deste tipo. O Google Scholar foi uma das bases escolhidas por apresentar uma quantidade massiva de trabalhos, o que pode atenuar essa desvantagem.

4 Resultados

4.1 N-gram Viewer

De todas arquiteturas pesquisadas, diversas não puderam ser pesquisadas com seus nomes diretamente devido aos problemas citados na Seção 2.1, dentre dessas temos ARM, SPARC, MIPS, (Motorola)68000, Intel pré x86, x86, x64, SPUR, SOAR. As arquiteturas que não tiveram este problema incluem PowerPC, PA-RISC, RISC I, RISC II e RISC-V.

4.1.1 ARM

Alguns dos nomes citados não representam uma só arquitetura, mas sim uma família de arquiteturas, como o ARM. Nestes casos, podemos usar diretamente o nome das versões das arquiteturas que existem dentro dessas arquiteturas, por exemplo, podemos analisar todas as versões de ARM existentes, que vão do ARMv1 até o ARMv8. As 3 primeiras versões do ARM não resultaram em nenhum valor, porém as subsequentes deram resultados significantes, com um aumento do uso a cada versão que foi lançada, que pode ser visto na Figura 2.

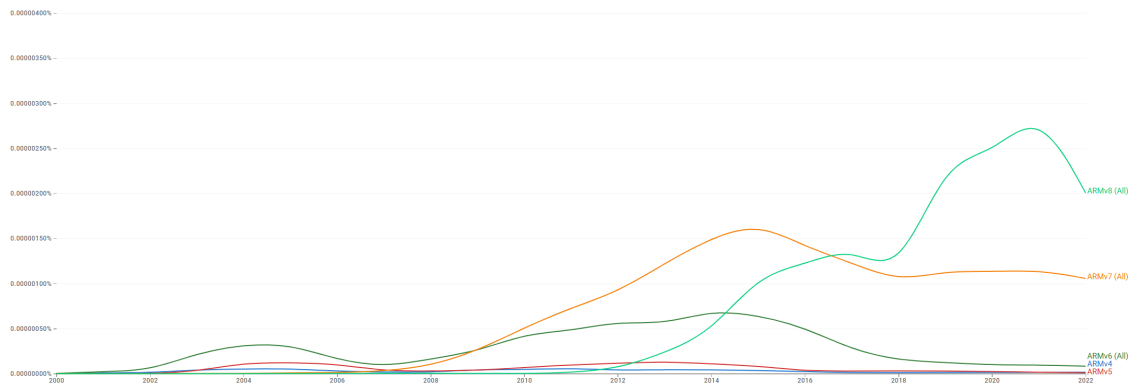


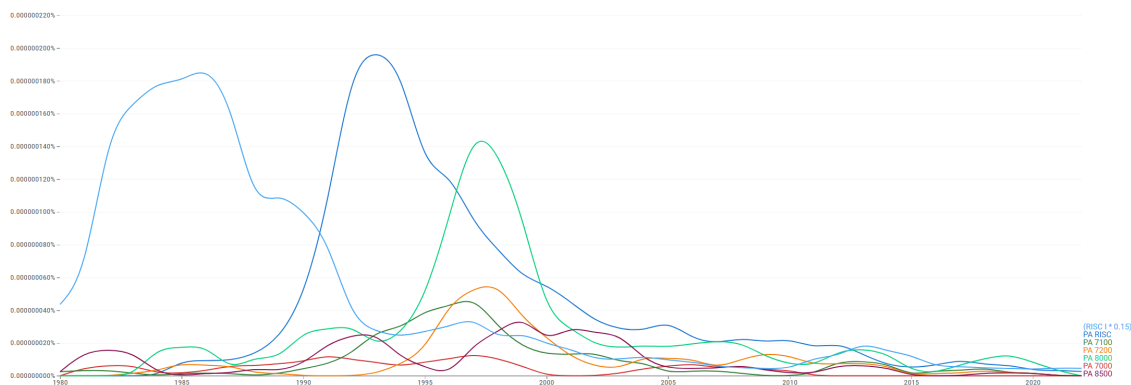
Figura 2 – Análise das versões do ARM

4.1.2 PA-RISC

O PA-RISC é uma arquitetura que pode ser pesquisada diretamente pelo nome. No entanto, também foi testada a inclusão de arquiteturas individuais na análise. O resultado, presente na Figura 3, apresentou valores baixos. Para comparação, foi adicionado o RISC I com um fator de 0,15x à sua frequência, considerando que o RISC I já possui uma frequência baixa, como será mostrado posteriormente.

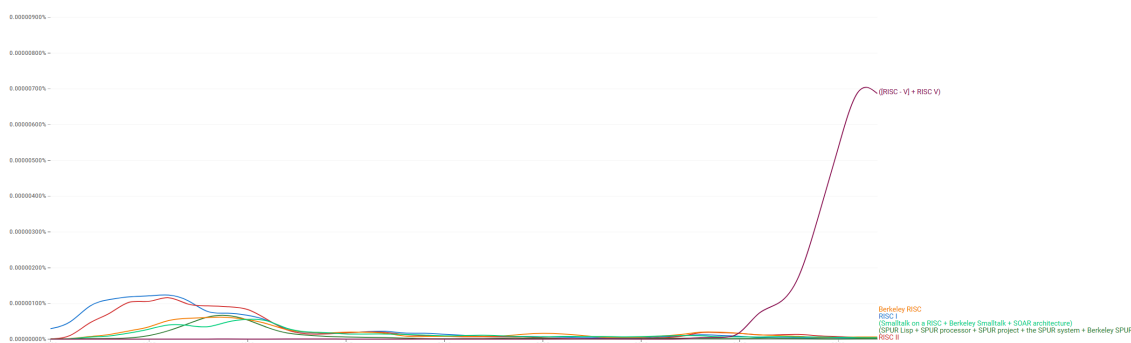
4.1.3 Berkeley RISC

A Figura 4 mostra uma comparação entre as arquiteturas Berkeley RISC. Exceto o RISC-V, essas arquiteturas surgiram como projetos acadêmicos e não possuem grande



frequência de uso por não serem produtos comerciais. No entanto, ainda é importante analisar essas arquiteturas, especialmente as duas primeiras, que cunharam o termo RISC.

O RISC III e o RISC IV não eram chamados assim quando foram publicados. Em vez disso, foram chamados de SOAR (Smalltalk On A RISC) (SAMPLES; KLEIN; FOLEY, 1985) e SPUR (Symbolic Processing Using RISCs) (HILL et al., 1985), respectivamente. Apesar disso, o professor que orientou a criação dos RISCs originais, Patterson, considera esses como os RISCs III e IV (PATTERSON; CHEN, 2016).



Tanto o SOAR quando o SPUR não puderam ser pesquisados diretamente, pois estes termos são utilizados em outros trabalhos não relacionados, e para solucionar isso, foram utilizados n-grams alternativos ou combinações que filtrassem resultados indesejados, e estes foram adicionados para obtenção de um gráfico único por arquitetura. Os resultados obtidos confirmam o fato de que as 3^o e 4^o versões não foram tão conhecidas como as duas primeiras, porém ainda há uma relevância significativa.

O RISC-V inicialmente também foi um projeto inicialmente desenvolvido na Universidade de Berkeley, porém este se tornou um padrão aberto da indústria, comandado pela Fundação RISC-V(PATTERSON; CHEN, 2016), e como visto na Figura 4, esta versão tem se tornado bem mais popular que as suas anteriores.

4.1.4 MIPS

A arquitetura MIPS ([MIPS Tech LLC, 2016](#)), assim como o ARM, foi pesquisada pelo nome de suas versões, conforme mostrado na Figura 5. Embora tenha perdido popularidade nos últimos anos, ela possui características similares às do RISC.

Assim como a arquitetura Berkeley RISC, a arquitetura MIPS também surgiu de uma universidade e se tornou uma arquitetura comercial através da empresa MIPS Technologies, pouco após as primeiras publicações do RISC ([PATTERSON; HENNESSY, 1996](#)). Ela tem sido utilizada em algumas universidades para ensinar conceitos sobre arquitetura e organização de computadores ([VOLLMAR; SANDERSON, 2005](#)), mostrando o seu valor acadêmico.

Segundo [Turley \(2021\)](#), a MIPS Technologies decidiu abandonar sua arquitetura tradicional e adotar a arquitetura RISC-V. Após passar por várias mudanças de propriedade e sair da falência, a empresa agora se concentra em desenvolver chips baseados na arquitetura RISC-V. Esta mudança marca o fim da linha para a família de CPUs MIPS e representa uma transformação completa no modelo de negócios da empresa, que antes se dedicava ao licenciamento de IP.

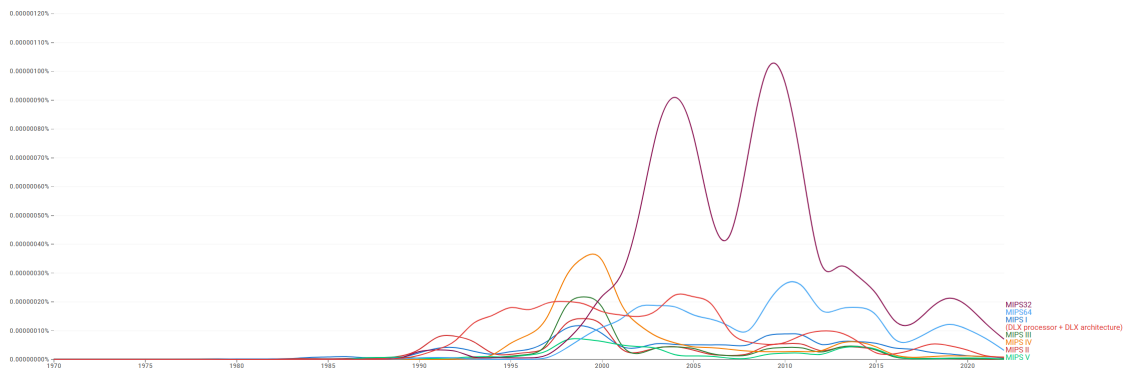


Figura 5 – Análise do MIPS

A DLX ([PATTERSON; HENNESSY, 1996](#)), uma arquitetura didática com diversas inspirações, sendo a principal delas a MIPS, foi incluída para comparação. Ela apresentou uma frequência de uso mais alta do que o esperado, mesmo que apenas nos casos onde acompanhava os termos "processor" e "architecture".

4.1.5 Intel Pré x86

As arquiteturas da Intel até o x86 foram analisadas, todas com o nome Intel junto do número. Apesar destas arquiteturas serem mais antigas em comparação a outras na pesquisa, elas ainda podem apresentar algum valor histórico ou características de interesse.

4.1.6 Intel Pós x86

As arquiteturas x86 e x64 são as mais comuns nos computadores domésticos atuais, sendo vendidas pelas empresas Intel e AMD, conforme visto em sites como [UserBenchmark](#) e [Steam Hardware Survey](#).

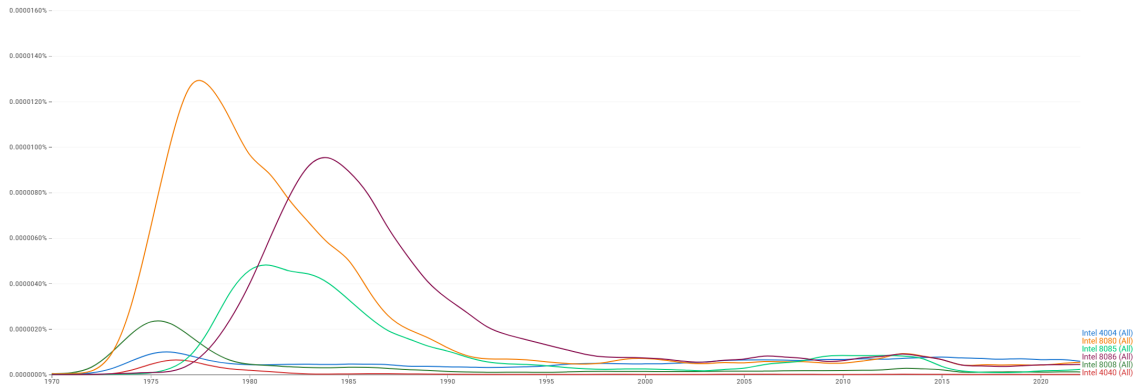


Figura 6 – Análise de arquiteturas Intel pré x86

Alguns computadores possuem processadores ARM, mas esses ainda são a minoria fora do mercado de dispositivos móveis.

Uma análise sobre a frequência de uso dessas arquiteturas é um pouco difícil, pois os termos x86 e x64 trazem algum ruído por serem utilizados em outros contextos. Apesar disso, o resultado ainda é interessante, pois apesar da defasagem do x86 (32 bits) em favor do x64 (64 bits) nos últimos anos, sua frequência de uso na literatura tem se mantido bem mais alta que seu sucessor.

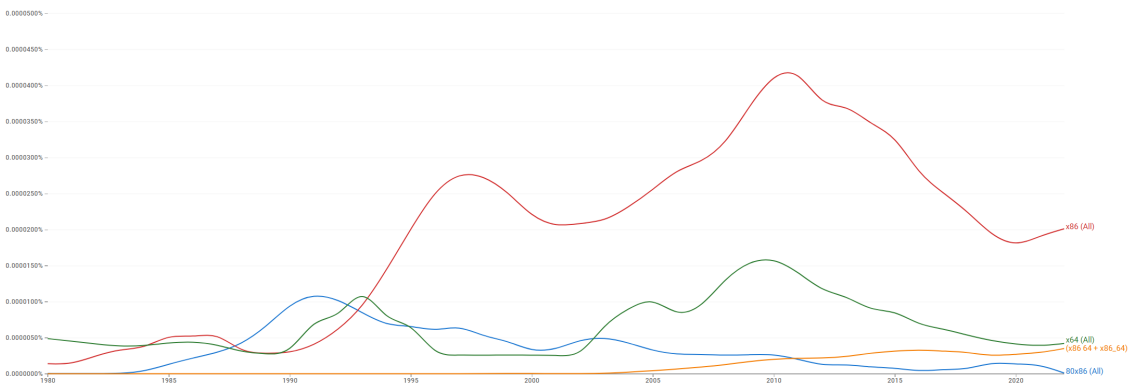


Figura 7 – Análise de arquiteturas pós x86

4.1.7 PowerPC

A arquitetura PowerPC surgiu da aliança entre Apple, Motorola e IBM, e apesar de ter seu uso defasado durante anos, esta teve um grande pico de uso durante os anos 90 como visto na Figura 8, sendo utilizada até o início dos anos 2000.

4.1.8 SPARC

A arquitetura SPARC, criada pela Sun Microsystems, é interessante por ser uma derivação direta do RISC-II (PATTERSON; HENNESSY, 1996, p. 116). A pesquisa por este termo revelou uma frequência de uso anterior ao seu lançamento, como pode ser visto

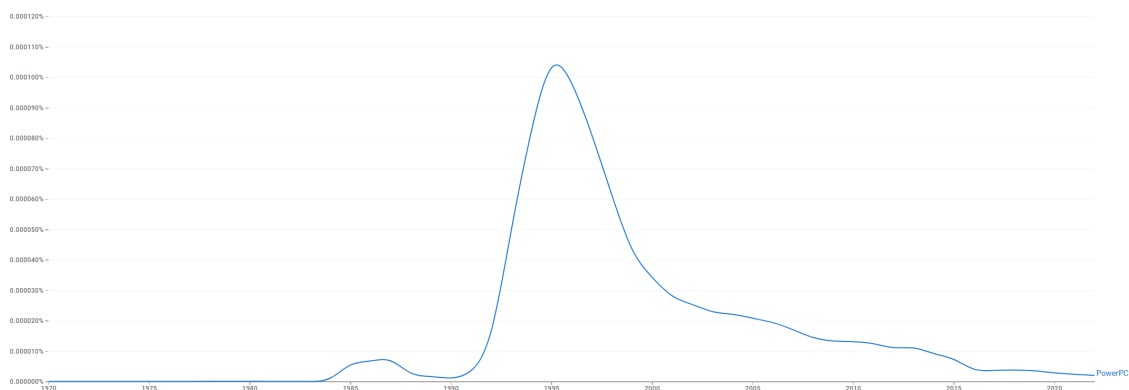


Figura 8 – Análise do PowerPC

na Figura 9. Esta figura também mostra o uso do recurso de *wildcards*, que foi utilizado para identificar quais palavras acompanham o termo principal mais frequentemente, auxiliando na filtragem em alguns casos.

Embora seja útil em certas situações, no caso do SPARC, este recurso de *wildcards* não foi tão eficaz, pois os termos mais comuns que acompanham o SPARC eram palavras conectivas, que não ofereciam o contexto do uso.

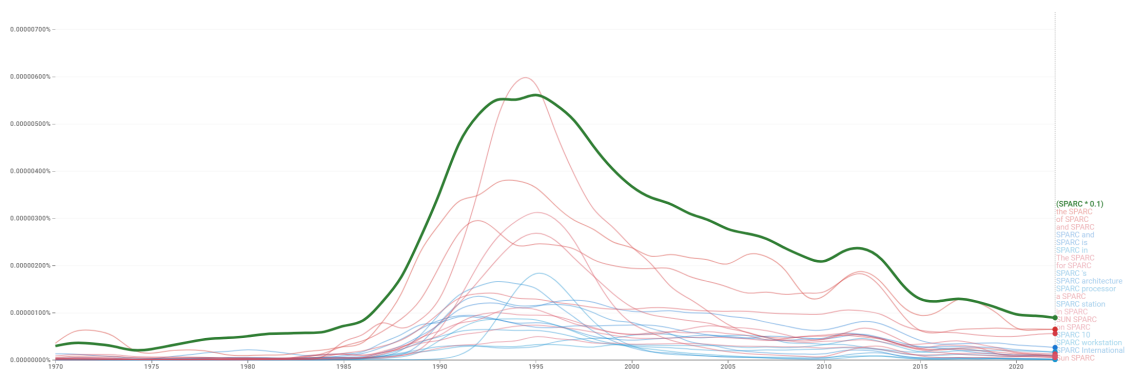


Figura 9 – Análise do SPARC

O Ngram Viewer permite pesquisar as obras que utilizaram os termos, filtradas por época. Este recurso foi aproveitado para analisar os contextos em que o termo SPARC foi utilizado. Os contextos encontrados foram X3 SPARC, ANSI SPARC, SPARC DBMS e SPARC system, permitindo a subtração destes do termo SPARC. Para esta subtração, esses termos foram multiplicados por um fator de 10, o que reduziu a quantidade no gráfico para aproximadamente 0% durante o período que a arquitetura SPARC ainda não existia.

Embora este processo introduza algumas imprecisões, o resultado, visto na Figura 10, ainda possui relevância, indicando de forma mais clara o período em que o termo começou a se popularizar.

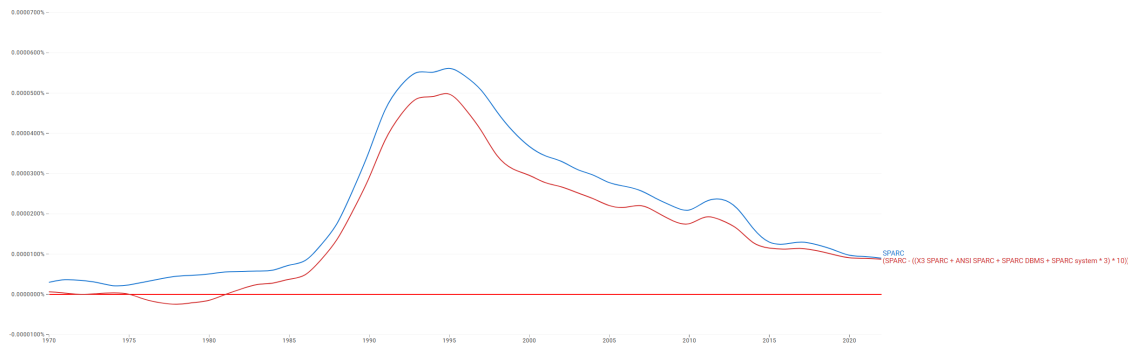


Figura 10 – Análise filtrada do SPARC

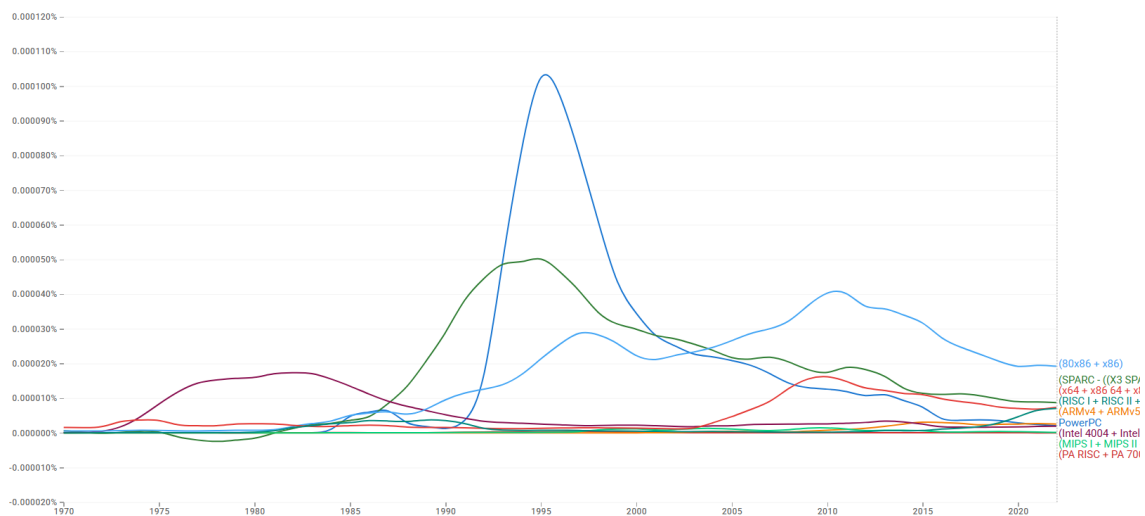


Figura 11 – Comparação entre todas as arquiteturas analisadas

4.1.9 Análise Geral

Ao analisar todos os *n-grams* em conjunto (Figura 11), observa-se que o PowerPC possui a maior frequência de uso, seguido pelas arquiteturas SPARC e x86. No entanto, exceto pela x86, essas arquiteturas têm perdido relevância ao longo dos anos, como indicado pela queda na frequência de uso dos termos na literatura.

Para uma análise dos termos menos utilizados, foi criado um gráfico excluindo os *n-grams* mais frequentes, presente na figura 11. Ao analisar todas as arquiteturas, observa-se que o Berkeley RISC (atualmente RISC-V) destaca-se como a última arquitetura com um rápido aumento na frequência de uso, similar ao de outras arquiteturas que se tornaram populares em algum momento. Isso indica que ele pode ser um bom tema para estudos mais aprofundados.

Outra arquitetura interessante para estudo é a família ARM. Como mencionado anteriormente, o termo ARM não foi utilizado por certos fatores, o que diminui o seu gráfico de uso, porém, esta é uma arquitetura que está presente em 99% dos *smartphones*, além de compor cerca de 50% de todos os processadores utilizados atualmente(ARM,).

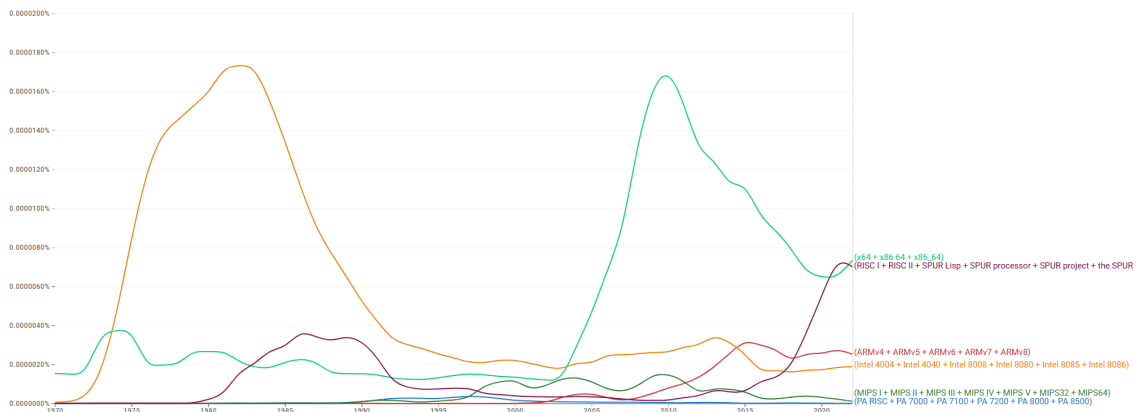


Figura 12 – Comparação entre as arquiteturas com menor frequência

4.2 Análise em Bases Acadêmicas

Foram realizadas oito pesquisas para cada arquitetura utilizando o Google Scholar: uma com o nome da arquitetura, denominada "Termo Individual", e sete combinando o nome da arquitetura com algum termo adicional para melhor filtragem dos resultados. A Tabela 1 mostra os resultados dessas pesquisas, agrupados em quatro partes: o Termo Individual, os quatro termos com melhores resultados, a média e o desvio padrão entre eles, e, por fim, os três termos menos relevantes.

Os termos com os melhores resultados incluem palavras que geralmente são esperadas em trabalhos relacionados a arquiteturas computacionais, como *architecture* (arquitetura), *processor* (processador), *instruction set* (conjunto de instruções) e *computing* (computação). Ao realizar as buscas, o nome da arquitetura foi mantido entre aspas para correspondências exatas, porém o termo adicional não, o que permitiu uma filtragem mais ampla. Por exemplo, ao invés de aceitar apenas resultados com o termo *computing*, o Google Scholar também aceita resultados com o termo *computer* (computador), que, apesar de ser uma palavra diferente, ainda se encaixa no contexto desejado.

Os valores dos melhores resultados têm seu desempenho e precisão medidos pela média e desvio padrão. Espera-se que esta média represente um resultado mais filtrado em comparação ao Termo Individual, que pode estar presente em situações não relacionadas ao que se deseja, similar ao que ocorreu com as buscas no Ngram Viewer. Esta média pode indicar a relevância real de um resultado em comparação aos outros. O desvio padrão, por outro lado, é uma métrica do quanto esses termos podem afetar a quantidade de resultados, e, neste caso, quanto menor, melhor, indicando que, independentemente do termo escolhido, a quantidade de resultados é sempre consistente.

Entre os termos menos relevantes, dois são mais focados em aplicações específicas - *embedded system* (sistema embarcado) e *SoC* (Sistema em um Chip) - e um é menos utilizado - *micro architecture* (microarquitetura) - que acaba sendo um subconjunto de arquitetura. Estes termos retornaram menos resultados no geral, mostrando que eles podem não ser ideais para pesquisas gerais sobre arquiteturas. Isso fica claro observando a Tabela 2, que mostra que a porcentagem de resultados encontrados em relação ao nome da arquitetura isolado é bem menor em comparação ao resto dos termos, com o valor máximo de 65,6%.

| Arquitetura | Termo Individual | Architecture | Processor | Instruction Set | Computing | # | Média | Desvio Padrão | Embedded System | SoC | Microarchitecture |
|-----------------|------------------|--------------|-----------|-----------------|-----------|---------|---------|---------------|-----------------|---------|-------------------|
| ARM | 5850000 | 3240000 | 1630000 | 2180000 | 4340000 | 2847500 | 1198537 | | 3090000 | 2660000 | 42800 |
| MIPS | 649000 | 99000 | 94300 | 106000 | 208000 | 126825 | 54330 | | 119000 | 60900 | 18800 |
| x86 | 171000 | 117000 | 109000 | 84200 | 129000 | 109800 | 18943 | | 68500 | 30700 | 22400 |
| SPARC | 261000 | 76000 | 62600 | 58500 | 113000 | 77525 | 24803 | | 61900 | 64200 | 9100 |
| PowerPC | 92300 | 61400 | 67700 | 55100 | 74200 | 64600 | 8211 | | 45900 | 23000 | 8290 |
| x64 | 74700 | 25000 | 21600 | 15500 | 36500 | 24650 | 8824 | | 15300 | 18200 | 1130 |
| RISC-V | 26400 | 22200 | 22200 | 19000 | 22100 | 21375 | 1584 | | 15700 | 11700 | 6420 |
| Intel(pré x86') | 96300 | 16800 | 4850 | 18600 | 16800 | 14263 | 6332 | | 6380 | 22700 | 70 |
| ARMv8 | 13000 | 10200 | 9190 | 7300 | 12000 | 9673 | 1963 | | 7740 | 7260 | 2670 |
| DLX | 49300 | 8830 | 3590 | 7220 | 17500 | 9285 | 5899 | | 8620 | 16600 | 987 |
| PA-RISC | 8870 | 7970 | 8330 | 7110 | 8180 | 7898 | 545 | | 3020 | 1820 | 4110 |
| RISC I | 11800 | 1330 | 1020 | 1160 | 1990 | 1375 | 429 | | 572 | 1410 | 275 |
| Berkeley RISC | 723 | 692 | 706 | 653 | 699 | 688 | 24 | | 334 | 259 | 474 |
| RISC II | 994 | 642 | 585 | 575 | 663 | 616 | 43 | | 219 | 472 | 181 |

Tabela 1 – Quantidade de Resultados de Pesquisa no Google Scholar

A Tabela 2 complementa os resultados apresentados na Tabela 1, convertendo a quantidade de resultados em percentuais relativos aos resultados obtidos utilizando apenas o Termo Individual. Com exceção do desvio padrão, observa-se que uma menor porcentagem indica resultados menos satisfatórios ao utilizar somente o nome da arquitetura, pois isso sugere que o nome está sendo empregado em contextos não relacionados aos termos adicionais. Além disso, esta tabela facilita a análise do desvio padrão, uma vez que a utilização de percentuais permite uma comparação mais precisa entre as diferentes arquiteturas.

Ambas tabelas foram ordenadas a partir da média, o que resulta em diferentes perspectivas. Na primeira, quanto maior a média de resultados com o parâmetro adicional, mais popular e relevante podemos considerar esta arquitetura. Por outro lado na Tabela 2, como dito anteriormente, quanto menor o percentual, menos o nome da arquitetura é utilizado no contexto de computação. O RISC I por exemplo foi encontrado em diversos contextos médicos que não se encaixam com o desejado.

| Arquitetura | Termo Individual | % Architect | % Process | % Instruction | % Comput | Média | Desvio Padrão | % Embedded Syst | % SoC | % Microarchitectur |
|-----------------|------------------|-------------|-----------|---------------|----------|-------|---------------|-----------------|-------|--------------------|
| Berkeley RISC | 723 | 95,7% | 97,6% | 90,3% | 96,7% | 95,1% | 3,3% | 46,2% | 35,8% | 65,6% |
| PA-RISC | 8870 | 89,9% | 93,9% | 80,2% | 92,2% | 89,0% | 6,1% | 34,0% | 20,5% | 46,3% |
| RISC-V | 26400 | 84,1% | 84,1% | 72,0% | 83,7% | 81,0% | 6,0% | 59,5% | 44,3% | 24,3% |
| ARMv8 | 13000 | 78,5% | 70,7% | 56,2% | 92,3% | 74,4% | 15,1% | 59,5% | 55,8% | 20,5% |
| PowerPC | 92300 | 66,5% | 73,3% | 59,7% | 80,4% | 70,0% | 8,9% | 49,7% | 24,9% | 9,0% |
| x86 | 171000 | 68,4% | 63,7% | 49,2% | 75,4% | 64,2% | 11,1% | 40,1% | 18,0% | 13,1% |
| RISC II | 994 | 64,6% | 58,9% | 57,8% | 66,7% | 62,0% | 4,3% | 22,0% | 47,5% | 18,2% |
| ARM | 5850000 | 55,4% | 27,9% | 37,3% | 74,2% | 48,7% | 20,5% | 52,8% | 45,5% | 0,7% |
| x64 | 74700 | 33,5% | 28,9% | 20,7% | 48,9% | 33,0% | 11,8% | 20,5% | 24,4% | 1,5% |
| SPARC | 261000 | 29,1% | 24,0% | 22,4% | 43,3% | 29,7% | 9,5% | 23,7% | 24,6% | 3,5% |
| MIPS | 649000 | 15,3% | 14,5% | 16,3% | 32,0% | 19,5% | 8,4% | 18,3% | 9,4% | 2,9% |
| DLX | 49300 | 17,9% | 7,3% | 14,6% | 35,5% | 18,8% | 12,0% | 17,5% | 33,7% | 2,0% |
| Intel(pré x86') | 96300 | 17,4% | 5,0% | 19,3% | 17,4% | 14,8% | 6,6% | 6,6% | 23,6% | 0,1% |
| RISC I | 11800 | 11,3% | 8,6% | 9,8% | 16,9% | 11,7% | 3,6% | 4,8% | 11,9% | 2,3% |

Tabela 2 – Comparação de Proximidade dos Resultados de Pesquisa no Google Scholar

| Arquitetura ▾ | Termo Individu. ▾ | Architectur ▾ | Processor ▾ | Instruction ▾ | Computing ▾ | # Média ▾ | Desvio Pad ▾ |
|-----------------|-------------------|---------------|-------------|---------------|-------------|-----------|--------------|
| ARM | 535912 | 10829 | 2839 | 1367 | 8287 | 5831 | 4468 |
| MIPS | 42944 | 1072 | 621 | 345 | 1351 | 847 | 450 |
| x86 | 2060 | 1014 | 639 | 244 | 1253 | 788 | 442 |
| RISC-V | 900 | 556 | 494 | 284 | 593 | 482 | 138 |
| SPARC | 6645 | 397 | 249 | 101 | 435 | 296 | 152 |
| PowerPC | 665 | 314 | 353 | 59 | 381 | 277 | 148 |
| ARMv8 | 172 | 116 | 72 | 28 | 120 | 84 | 43 |
| DLX | 939 | 46 | 43 | 21 | 40 | 38 | 11 |
| x64 | 242 | 31 | 33 | 11 | 52 | 32 | 17 |
| PA-RISC | 32 | 25 | 15 | 15 | 28 | 21 | 7 |
| Intel(pré x86') | 273 | 27 | 15 | 4 | 32 | 20 | 13 |
| Berkeley RISC | 12 | 10 | 7 | 7 | 12 | 9 | 2 |
| RISC I | 30 | 8 | 4 | 5 | 7 | 6 | 2 |
| RISC II | 37 | 3 | 2 | 3 | 3 | 3 | 1 |

Tabela 3 – Quantidade de Resultados de Pesquisa no Periódicos Capes

Observa-se nas Tabelas 2 e 4 que os nomes mais distintos apresentaram os maiores percentuais consistentemente nas duas bases de dados analisadas. Entre esses nomes, destacam-se Berkeley RISC, PA-RISC, RISC-V, ARMv8, PowerPC e x86. Como mostrado na Tabela 5 Estes tiveram uma variação percentual entre 20,1% e 26% por arquitetura entre bases acadêmicas além de um desvio padrão entre 9,3% e 17,1% menor na base de dados Periódicos CAPES, indicando resultados mais consistentes ao utilizar diferentes termos adicionais.

| Arquitetura ▾ | Termo Individu. ▾ | % Archite ▾ | % Proces ▾ | % Instruc ▾ | % Compl ▾ | % Média ▾ | % Desvio ▾ |
|-----------------|-------------------|-------------|------------|-------------|-----------|-----------|------------|
| Berkeley RISC | 12 | 83,3% | 58,3% | 58,3% | 100,0% | 75,0% | 20,4% |
| PA-RISC | 32 | 78,1% | 46,9% | 46,9% | 87,5% | 64,8% | 21,1% |
| RISC-V | 900 | 61,8% | 54,9% | 31,6% | 65,9% | 53,5% | 15,3% |
| ARMv8 | 172 | 67,4% | 41,9% | 16,3% | 69,8% | 48,8% | 25,1% |
| PowerPC | 665 | 47,2% | 53,1% | 8,9% | 57,3% | 41,6% | 22,2% |
| x86 | 2060 | 49,2% | 31,0% | 11,8% | 60,8% | 38,2% | 21,4% |
| RISC I | 30 | 26,7% | 13,3% | 16,7% | 23,3% | 20,0% | 6,1% |
| x64 | 242 | 12,8% | 13,6% | 4,5% | 21,5% | 13,1% | 6,9% |
| RISC II | 37 | 8,1% | 5,4% | 8,1% | 8,1% | 7,4% | 1,4% |
| Intel(pré x86') | 273 | 9,9% | 5,5% | 1,5% | 11,7% | 7,1% | 4,6% |
| SPARC | 6645 | 6,0% | 3,7% | 1,5% | 6,5% | 4,4% | 2,3% |
| DLX | 939 | 4,9% | 4,6% | 2,2% | 4,3% | 4,0% | 1,2% |
| MIPS | 42944 | 2,5% | 1,4% | 0,8% | 3,1% | 2,0% | 1,0% |
| ARM | 535912 | 2,0% | 0,5% | 0,3% | 1,5% | 1,1% | 0,8% |

Tabela 4 – Comparação de Proximidade dos Resultados de Pesquisa no Periódicos Capes

| Coluna 1 ▾ | % Média Scholar ▾ | % Média Capes ▾ | % Diferença Média ▾ | Desvio Scholar ▾ | Desvio Capes ▾ | Diferença Desvio ▾ |
|-----------------|-------------------|-----------------|---------------------|------------------|----------------|--------------------|
| Berkeley RISC | 95,1% | 75,0% | 20,1% | 3,3% | 20,4% | -17,1% |
| PA-RISC | 89,0% | 64,8% | 24,2% | 6,1% | 21,1% | -15,0% |
| RISC-V | 81,0% | 53,5% | 27,4% | 6,0% | 15,3% | -9,3% |
| ARMv8 | 74,4% | 48,8% | 25,6% | 15,1% | 25,1% | -10,0% |
| PowerPC | 70,0% | 41,6% | 28,4% | 8,9% | 22,2% | -13,3% |
| x86 | 64,2% | 38,2% | 26,0% | 11,1% | 21,4% | -10,4% |
| RISC II | 62,0% | 7,4% | 54,6% | 4,3% | 1,4% | 3,0% |
| ARM | 48,7% | 1,1% | 47,6% | 20,5% | 0,8% | 19,7% |
| x64 | 33,0% | 13,1% | 19,9% | 11,8% | 6,9% | 4,9% |
| SPARC | 29,7% | 4,4% | 25,3% | 9,5% | 2,3% | 7,2% |
| MIPS | 19,5% | 2,0% | 17,6% | 8,4% | 1,0% | 7,3% |
| DLX | 18,8% | 4,0% | 14,8% | 12,0% | 1,2% | 10,8% |
| Intel(pré x86') | 14,8% | 7,1% | 7,7% | 6,6% | 4,6% | 2,0% |
| RISC I | 11,7% | 20,0% | -8,3% | 3,6% | 6,1% | -2,4% |

Tabela 5 – Diferença da média e do desvio padrão entre as bases acadêmicas.

Por outro lado, as arquiteturas com as menores médias percentuais apresentaram uma variação muito mais ampla, com RISC II e ARM retornando, respectivamente, aproximadamente 54,6% e 47,6% de resultados a menos na base de dados Periódicos CAPES. De modo geral, o desvio padrão dessas arquiteturas no CAPES foi menor, o que pode ser atribuído à alta discrepância na média de resultados relevantes entre as bases de dados.

4.3 Análise Geral

Ao comparar os resultados obtidos nas pesquisas com os resultados dos Ngrams, podemos corroborar a precisão de certos achados e corrigir outros. A arquitetura ARM, apesar de sua relevância significativa na atualidade, não apresentou altos valores nos Ngrams devido à filtragem realizada, que considerou apenas os resultados das arquiteturas individuais. Nas pesquisas, foram utilizados tanto os termos "ARM" quanto "ARMv8", uma das versões do ARM. O nome genérico ARM, como esperado, possuía uma grande quantidade de ruído nos resultados, o que foi parcialmente mitigado pela filtragem com termos adicionais, embora não o suficiente para removê-lo da primeira posição.

Por outro lado, a versão individual "ARMv8", assim como nos Ngrams, não apresentou tantos resultados quanto o nome genérico da arquitetura. No entanto, obteve uma alta média percentual de resultados relevantes, demonstrando que, apesar de ser uma filtragem rigorosa, ela aumenta significativamente a precisão dos resultados.

De forma semelhante ao observado no Ngram Viewer, a arquitetura x86 destacou-se com um dos maiores resultados, enquanto sua sucessora, x64, posicionou-se duas posições abaixo no Google Scholar, confirmando sua relevância. Em contraste, a arquitetura SPARC, que possuía uma frequência mais alta no Ngram Viewer, ficou abaixo do x86. Este dado sugere duas possibilidades: ou o nome x86 foi utilizado com menos frequência em comparação ao SPARC, apesar da maior quantidade de documentos, ou a filtragem no Ngram Viewer não foi suficientemente eficaz. Independentemente disso, o fato de ambas as arquiteturas estarem próximas nas duas métricas confirma um certo nível de consistência nos dados.

Apesar da arquitetura MIPS ter alcançado valores elevados, pesquisas pelas suas versões individuais resultaram em valores baixos, em torno de 2000 resultados, similar à frequência dos seus Ngrams. Isso pode ser devido ao fato de que o termo MIPS pode se referir a "Million Instructions Per Second" (Milhões de Instruções por Segundo), uma métrica usada para medir a velocidade de um processador. Logo, mesmo que a quantidade de resultados seja alta, há uma chance de que parte desses resultados contenha ruído dentro da própria área de computação.

O PowerPC, como esperado, foi uma das arquiteturas com menos ruído e, como indicado pelos Ngrams, a quantidade de resultados foi relativamente alta, embora ainda menor que a do x86 e do SPARC. Os Ngrams mostram que, apesar do PowerPC possuir o maior pico de todos, este não durou muito tempo, com um período de pouco menos de uma década, o que pode explicar sua posição nas tabelas.

O RISC-V obteve o terceiro maior percentual de resultados relevantes. Apesar de ser uma arquitetura recente (2010), retornou uma quantidade significativa de resultados, não muito distante do x64, que surgiu mais de uma década antes (1999). O rápido aumento da frequência de uso exibido nos Ngrams e a quantidade de resultados indicam que o RISC-V pode ser um forte competidor contra outras arquiteturas recentes.

A arquitetura PA-RISC apresentou mais resultados do que o esperado, considerando os dados dos Ngrams, além de ter o segundo maior percentual de fontes relevantes em ambas as bases acadêmicas. Apesar disso, ainda não possui grande popularidade, sendo uma arquitetura comercial. Para comparação, o PA-RISC ficou atrás da arquitetura DLX, que é puramente acadêmica.

As arquiteturas RISC originais, apesar de sua importância e influência nas arquiteturas RISC atuais, resultaram em uma quantidade ínfima de resultados em comparação com as outras, ficando nas três últimas posições na tabela. O termo Berkeley RISC retornou

o maior percentual de resultados relevantes entre todas as arquiteturas, contrastando com as arquiteturas RISC I e RISC II, que apresentaram grande ruído devido à utilização desses termos em pesquisas médicas recentes.

Considerações finais

Embora os métodos de pesquisa aplicados neste estudo tenham fornecido resultados satisfatórios para trabalhos onde a precisão não é crucial, em contextos que exigem alta precisão, seria necessário um exame mais detalhado. A filtragem no Google Ngram Viewer deve ser realizada com maior cautela, analisando minuciosamente os termos que causam ruído nos resultados. Esse método funciona bem para análises de relevância ao longo do tempo, mostrando os períodos em que determinados termos foram mais utilizados.

A aplicação conjunta de ambos os métodos de pesquisa ajudou a mitigar suas limitações, pois cada método complementou as informações que o outro não apresentava. Enquanto o Visualizador de Ngrams não oferece uma métrica eficiente de relevância geral entre termos, as pesquisas em bases acadêmicas não permitem uma análise temporal de relevância de forma fácil. Com ambos os métodos, podemos determinar se uma arquitetura ainda é relevante e qual é seu nível de relevância em comparação com outras.

Para pesquisas futuras, recomenda-se o desenvolvimento de métodos de filtragem mais robustos para melhorar a precisão dos resultados. Além disso, é essencial realizar uma pesquisa mais minuciosa que envolva as diversas arquiteturas individuais, em vez de termos genéricos como "ARM" em todos métodos envolvidos. Esta abordagem mais detalhada não foi adotada neste estudo devido a limitações de tempo e à falta de necessidade deste nível de detalhamento no contexto atual.

A Webbibliometry About the Relevance of Computer Architectures

Diogo Valadares Reis dos Santos*, David Vasconcelos Corrêa da Silva

2025

Abstract

This article explores the educational application of computational architectures by conducting a bibliometric analysis to evaluate the relevance of RISC architectures within the broader context of computing systems. Leveraging tools such as the Google Books Ngram Viewer and academic databases like Google Scholar and CAPES Journals, the study identifies trends and patterns in the evolution and adoption of computational architectures. Key findings include the chronological prominence of RISC-based designs and their influence on modern architectures such as RISC-V. The research combines historical usage data with an evaluation of academic publication trends, offering insights into the enduring significance of these architectures for educational and technological advancements.

Key-words: computational architectures, RISC, bibliometric analysis, Ngram Viewer, academic relevance.

Referências

ARM. *Building the Future of Computing*. Disponível em: <<https://www.arm.com/>>. Citado na página 10.

COSTA, H. Modelo para webibliomining: proposta e caso de aplicação. v. 13, p. 115–126, 2010. Citado 2 vezes nas páginas 1 e 2.

HILL, M. D. et al. *SPUR: A VLSI Multiprocessor Workstation*. Berkeley, 1985. Citado na página 6.

*diogo-valadares@hotmail.com

LIN, Y. et al. Syntactic Annotations for the Google Books NGram Corpus. jul. 2012. Citado na página 3.

MICHEL, J.-B. et al. Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, v. 331, n. 6014, p. 176–182, jan. 2011. ISSN 0036-8075, 1095-9203. Disponível em: <<https://www.science.org/doi/10.1126/science.1199644>>. Citado 3 vezes nas páginas 1, 2 e 3.

MIPS Tech LLC. *MIPS32 Architecture For Programmers*. 6. ed. [S.l.: s.n.], 2016. v. 1-2. Citado na página 7.

PATTERSON, D. A.; CHEN, T. *RISC-V Geneology*. Berkeley, 2016. 8 p. Disponível em: <<https://www2.eecs.berkeley.edu/Pubs/TechRpts/2016/EECS-2016-6.pdf>>. Citado 2 vezes nas páginas 2 e 6.

PATTERSON, D. A.; HENNESSY, J. L. *Computer architecture : a quantitative approach*. 2nd ed. ed. San Francisco: Morgan Kaufmann Publishers, 1996. ISBN 978-1-55860-329-5 978-1-55860-372-1. Disponível em: <<http://archive.org/details/computerarchitec00patt>>. Citado 2 vezes nas páginas 7 e 8.

SAMPLES, A. D.; KLEIN, M.; FOLEY, P. *SOAR Architecture*. Berkeley, 1985. Disponível em: <<https://www2.eecs.berkeley.edu/Pubs/TechRpts/1985/5940.html>>. Citado na página 6.

TURLEY, J. *Wait, What? MIPS Becomes RISC-V*. 2021. Disponível em: <<https://www.eejournal.com/article/wait-what-mips-becomes-risc-v/>>. Citado na página 7.

VOLLMAR, K.; SANDERSON, P. A MIPS assembly language simulator designed for education. *Journal of Computing Sciences in Colleges*, jan. 2005. Disponível em: <<https://bearworks.missouristate.edu/articles-cnas/1039>>. Citado na página 7.