Supplementary Material



**Supplementary Figure 1:** Finding the optimal number of dimensions for dimensionality reduction. Here we show the test data Mean Squared reconstruction Error (MSE) after applying SVD. The smaller error indicates a more accurate reconstruction.



**Supplementary Figure 2:** Finding the optimal number of clusters using silhouette and Davies- Bouldin scores. Here we have normalized both scores between 0 and 1 with higher values indicating better clustering.

**Supplementary Table 1**: 74 work or professional titles included in the study picked from Statistics Finland table (<https://www2.tilastokeskus.fi/en/luokitukset/ammatti>). The table contains both the original Finnish terms and their corresponding English (non-official) translations.

|  |  |
| --- | --- |
| **Finnish** | **English** |
| Tieto- ja viestintäteknologiajohtajat | Information and Communication Technology Directors |
| Tieto- ja viestintäteknologiajohtaja | Information and Communication Technology Director |
| Johtaja, tieto- ja viestintäteknologia | Director of Information and Communication Technology |
| Tietoliikennesuunnittelija | Telecommunications Planner |
| ICT-alan erityisasiantuntijat | ICT Sector Specialists |
| Tietotekniikan insinööri | Information Technology Engineer |
| Tietoliikenneinsinööri | Telecommunications Engineer |
| Tieto- ja viestintätekniikan myynnin er. | ICT Sales Specialist |
| Erityisasiantuntija, tieto-, viestintätekn. myynti | Specialist in IT and Communication Technology Sales |
| Tieto- ja viestintäteknologian erityisasiantuntija | ICT Specialist |
| Systeemityön erityisasiantuntijat | System Work Specialists |
| Sovellusarkkitehdit | Application Architects |
| Atk-päällikkö | IT Manager |
| Tietohallintopäällikkö | Information Management Director |
| Projektipäällikkö, tietotekniikka | Project Manager, Information Technology |
| Systeeminsuunnittelija | Systems Designer |
| ICT-liiketoiminta-analyytikko | ICT Business Analyst |
| ICT-järjestelmäarkkitehti | ICT Systems Architect |
| ICT-järjestelmäanalyytikko | ICT Systems Analyst |
| Muu sovellusarkkitehti | Other Application Architect |
| Sovellussuunnittelijat | Application Designers |
| Atk-suunnittelija | IT Designer |
| Ohjelmistosuunnittelija | Software Designer |
| Käyttöliittymäsuunnittelija | User Interface Designer |
| Sisältösuunnittelija | Content Designer |
| Web- ja multimediakehittäjät | Web and Multimedia Developers |
| Web-suunnittelija | Web Designer |
| Sovellusohjelmoijat | Application Programmers |
| Ohjelmoija, tietotekniikka | Programmer, Information Technology |
| Muut ohjelmisto- ja sovelluskehittäjät | Other Software and Application Developers |
| Muu ohjelmisto- , sovellustyön erityisasiantuntija | Other Software and Application Work Specialist |
| Tietokantojen,- verkkojen ja ohjelmistojen er. | Databases, Networks, and Software Specialist |
| Tietokantasuunnittelijat ja -vastaavat | Database Designers and Managers |
| Erityisasiantuntija, tietokannat | Database Specialist |
| Tietojärjestelmien ylläpitäjät | Information System Administrators |
| Järjestelmäpäällikkö | Systems Manager |
| Tietoverkkojen erityisasiantuntijat | Network Specialists |
| Tietoverkkoasiantuntija | Network Expert |
| Muut tietokanta- ja tietoverkkojen erityisasiantun | Other Database and Network Specialists |
| Tietoturvapäällikkö | Information Security Manager |
| Verkostonhoitaja | Network Manager |
| Sulautettujen järjestelmien turvallisuusasiantunt. | Embedded Systems Security Expert |
| Tietotekniikan teknikko | Information Technology Technician |
| Muu elektroniiikan ja tietotekniikan teknikko | Other Electronics and Information Technology Technician |
| Informaatio- ja tietoliikenneteknologian a. | Information and Telecommunications Technology Assistant |
| ICT-alan teknikot ja käyttäjätukihenkilöt | ICT Technicians and User Support Personnel |
| Käytön operaattorit | Operations Operators |
| Käyttöpäällikkö, tietotekniikka | IT Operations Manager |
| Käytönsuunnittelija, tietotekniikka | IT Operations Planner |
| Operaattori, tietotekniikka | Operator, Information Technology |
| Palvelinkeskuksen käytönvalvoja | Data Center Operator |
| Käytön tukihenkilöt | Support Staff |
| Atk-neuvoja | IT Advisor |
| Mikrotukihenkilö | Micro Support Personnel |
| Atk-yhteyshenkilö | IT Contact Person |
| Muu käytön tukihenkilö | Other Support Personnel |
| Tietoverkkoteknikot | Network Technicians |
| Tietoverkkoteknikko | Network Technician |
| Webmasterit ja -teknikot | Webmasters and Technicians |
| Webmaster | Webmaster |
| Teleliikenne- sekä radio- ja tv-teknikot | Telecommunications, Radio and TV Technicians |
| Lähetys- ja audiovisuaaliteknikot | Broadcasting and Audiovisual Technicians |
| AV-teknikko | AV Technician |
| Muu lähetys- tai audiovisuaaliteknikko | Other Broadcasting or Audiovisual Technician |
| Televiestinnän tekniset asiantuntijat | Telecommunications Technical Experts |
| Tekninen asiantuntija, televiestintä | Technical Expert, Telecommunications |
| Muu televiestinnän tekninen asiantuntija | Other Technical Expert in Telecommunications |
| Tallentaja, tietotekniikka | Recorder, Information Technology |
| Jälkikäsittelijä, tietotekniikka | Post-Processor, Information Technology |
| Elektroniikka-asentaja | Electronics Installer |
| Muu elektroniikka-, automaatiolaitteiden asentaja | Other Electronics, Automation Equipment Installer |
| Tieto- ja viestintäteknologian asentajat,korjaajat | Information and Communication Technology Installers and Repairers |
| Tietoliikenneasentaja | Telecommunications Installer |
| Tietotekniikka-asentaja | IT Installer |

**Supplementary Table 2:** Full Annif output for an example job advertisement. Here neural ensemble model “yso-en” was used without upper or lower limits for term count and score thresholds.

**Input:** “We are looking for (Senior) Embedded Linux Designer at our office in Espoo.

You will be responsible for designing embedded software for [company name hidden] power electronics products. Your role will include software design and integrating the software and hardware into the working entity in co-operation with other project team members. You will report to the Manager of SW team.

To succeed in this position, you should have the following qualifications

-B.Sc. or M.Sc. degree within Software Engineering or Electrical Engineering

-Experience in C/C++ programming, Python and Shell scripting in embedded Yocto Linux environment

-Understanding of electronic control systems hardware is a benefit

-Experience in communication protocols is a benefit

-Fluent in English, both written and spoken

-Proactive hands-on attitude

-Good documentation skills and analytical mind set

-Good team working and communication skills

We offer you an opportunity to develop your skills in software design and power electronics as a member of a team of professionals. You will be offered interesting projects and challenges in an international environment. The position is full time. The starting date is negotiable.”

**Output:**

|  |  |  |
| --- | --- | --- |
| **Rank** | **Term** | **Score** |
| 1 | computer programmes | 0.39095 |
| 2 | Linux | 0.20689 |
| 3 | software design | 0.18790 |
| 4 | programming | 0.17400 |
| 5 | operating systems | 0.14437 |
| 6 | software development | 0.11020 |
| 7 | ubiquitous computing | 0.10773 |
| 8 | power electronics | 0.10301 |
| 9 | software technology | 0.10113 |
| 10 | electronics | 0.09902 |
| 11 | electronic documents | 0.09115 |
| 12 | electrical engineering | 0.08457 |
| 13 | teamwork | 0.08060 |
| 14 | Espoo | 0.06872 |
| 15 | development (active) | 0.06296 |
| 16 | Into Work Project | 0.06270 |
| 17 | professional skills | 0.06222 |
| 18 | electrical engineers | 0.05936 |
| 19 | design (artistic creation) | 0.05689 |
| 20 | teams | 0.04883 |
| 21 | communication | 0.04304 |
| 22 | professional development | 0.03621 |
| 23 | know-how | 0.03494 |
| 24 | work communities | 0.03288 |
| 25 | software developers | 0.03139 |
| 26 | repertoires (arts) | 0.03082 |
| 27 | software engineering | 0.01885 |
| 28 | free software | 0.01553 |
| 29 | systems designers | 0.01401 |
| 30 | electrical designers | 0.01394 |
| 31 | electricity works | 0.01375 |
| 32 | control engineering | 0.00898 |
| 33 | competence inventory | 0.00868 |
| 34 | enterprises | 0.00866 |
| 35 | qualification | 0.00781 |
| 36 | spoken programmes and talkshows | 0.00755 |
| 37 | automation | 0.00704 |
| 38 | maintenance | 0.00575 |
| 39 | planning and design | 0.00403 |
| 40 | electronic documentation | 0.00336 |
| 41 | communications technology | 0.00318 |
| 42 | user-centeredness | 0.00225 |
| 43 | testing | 0.00204 |
| 44 | modifiability | 0.00093 |

**Supplementary Table 3:** Full Annif output for an example thesis abstract. Here neural ensemble model “yso-en” was used without upper or lower limits for term count and score thresholds.

**Input: “**In response to the contemporary challenge of employee dissatisfaction arising from repetitive and mundane tasks, this study explores the potential of Robotic Process Automation (RPA) to optimize individual tasks within organizational processes. Recognizing the obstacles of high costs, lengthy implementation timeframes and the complexity associated with deployment of traditional automation technologies, this study focuses on the feasibility of small-scale, single-process automation through RPA. The study involves a literature review and interviews with employees from different business sectors to gain insights, subsequently proceeding to simulated implementation of RPA. Findings reveal employee interest in RPA technology, though a common misconception persists that automation necessitates extensive planning by management. To challenge the misconception, two processes from different companies are identified and simulated for automation, showcasing the potential for employees to initiate small-scale automation. In conclusion, this study illustrates how RPA effectively enhances small-scale organizational processes through task automaton. By utilizing common tools (Office 365) and encouraging individual employees to delve into RPA, organizations can achieve increased efficiency, higher employee satisfaction, and cost-effectiveness. This study contributed insights for businesses seeking practical approaches to enhance organizational processes through RPA implementation.**”**

**Output:**

|  |  |  |
| --- | --- | --- |
| **Rank** | **Term** | **Score** |
| 1 | automation | 0.56649 |
| 2 | leadership (activity) | 0.37664 |
| 3 | enterprises | 0.36515 |
| 4 | organisations (systems) | 0.35332 |
| 5 | employees | 0.28158 |
| 6 | processes | 0.21784 |
| 7 | self-evaluation | 0.14583 |
| 8 | simulation | 0.12416 |
| 9 | robots | 0.12235 |
| 10 | staff | 0.11403 |
| 11 | efficiency (properties) | 0.10534 |
| 12 | development (active) | 0.08654 |
| 13 | automation technology | 0.08395 |
| 14 | automation systems | 0.06166 |
| 15 | robotics | 0.05664 |
| 16 | quality management | 0.04483 |
| 17 | automatons | 0.03552 |
| 18 | technology | 0.02991 |
| 19 | committing oneself | 0.02923 |
| 20 | quality awards | 0.02763 |
| 21 | human resource management | 0.02763 |
| 22 | leadership (properties) | 0.02354 |
| 23 | business | 0.02056 |
| 24 | organisational changes | 0.02025 |
| 25 | business employees | 0.01431 |
| 26 | simulators | 0.01021 |
| 27 | encouragement | 0.00976 |
| 28 | success | 0.00515 |
| 29 | PC Tools Utilities | 0.00239 |
| 30 | tradition | 0.00144 |
| 31 | control systems | 0.00034 |

**Supplementary Table 4**: 28 topic clusters and their manually assigned labels found in the theses abstract data set with 18254 theses from ICT-field. For each cluster, we show 10 terms nearest to each cluster’s centroid. Table contains total 176 unique terms.

|  |  |  |
| --- | --- | --- |
| **No.** | **Cluster Terms** | **Cluster Label** |
| **1.** | Software development, computer programmes, repertoires, testing, applications (computer programmes), programming, software engineering, testing methods, software technology, development (active) | Software engineering |
| **2.** | Data security, safety and security, data protection, cyber security, information networks, data systems, risk management, shielding, authentication, control (prevention) | Data security |
| **3.** | Automation, data systems, processes, enterprises, development (active), product development, final projects (education), planning and design, systems, improving | Product development & Design |
| **4.** | Web pages, websites, Internet, usability, online services, content management, planning and design, final projects (education), enterprises, realization (active) | Content creation & Management |
| **5.** | Technology, information technology, development (passive), innovations, technological development, product development, final projects (education), automation, development (active), applications (computer programmes) | Technological development |
| **6.** | Projects, project work, project management, development projects, project planning, planning and design, development (active), final projects (education), project leadership, enterprises | Project management |
| **7.** | Machine learning, artificial intelligence, algorithms, neural networks (information technology), robotics, technology, robots, learning, automation, deep learning | Artificial Intelligence |
| **8.** | Games, computer games, digital games, game design, game sector, game industry, playing (games and sports), mobile games, video games, game programming | Game design, Programming & Playing |
| **9.** | Working life, work, employees, know-how, final projects (education), enterprises, development (active), staff, job description, learning | Work life |
| **10.** | Mobile devices, cell phones, applications (computer programmes), mobile apps, operating systems, smartphones, Android, programming, mobile services, mobile communications services | Mobile devices, Apps & OS |
| **11.** | Programming, programming languages, databases, computer programmes, Java, applications (computer programmes), data systems, user interfaces, object-oriented programming, final projects (education) | Programming |
| **12.** | Measurement, measuring instruments (devices), measuring methods, testing, automation, devices, electrical engineering, electronics, product development, measuring instruments (indicators) | Measurements & Testing |
| **13.** | Marketing, electronic commerce, business, enterprises, electronic business, digital marketing, business operations, Internet, services, products | E-commerce & Digital marketing |
| **14.** | Modelling (creation related to information), Three-dimensional imaging, three-dimensionality, planning and design, computer programmes, visualisation, final projects (education), computer-aided design, computer graphics, games | Computer aided design |
| **15.** | Servers, Linux, operating systems, computers, computer programmes, data systems, installation, information technology, information networks, virtualisation | Servers & OS |
| **16.** | Online services, services, cloud services, Internet, information networks, data systems, enterprises, information technology, use, applications (computer programmes) | Internet & Cloud services |
| **17.** | Windows XP, Windows 7, Windows 95, Windows 98, Windows NT, Windows 8, operating systems, Windows 10, Windows, Windows Server | Windows OS |
| **18.** | User-centeredness, usability, users, user interfaces, planning and design, product development, evaluation, user experience, web pages, final projects (education) | UI designs |
| **19.** | Wireless data transmission, wireless communication, wireless local area networks, telecommunications technology, wireless technology, mobile communication networks, data communications networks, local area networks, data transfer, wireless networks | Wireless communication |
| **20.** | Customers, customer orientation, customer service, services, customership, enterprises, marketing, development (active), customer satisfaction, customer relationship management | Customer services |
| **21.** | HTML, web pages, applications (computer programmes), websites, ASP.NET, description languages, Internet, planning and design, PHP, narrative theory | Web programming |
| **22.** | Universities of applied sciences, students, learning, teaching and instruction, study, education and training, learning environment, final projects (education), online teaching, information technology | Education & Learning |
| **23.** | Organisations (systems), leadership (activity), development (active), enterprises, self-evaluation, processes, success, business, data systems, quality awards | Organizations & Leaderships |
| **24.** | Video, video recording, editing, video equipment, image processing, production, video tape recorders, films, final projects (education), cameras | Image & Video processing |
| **25.** | Data communications networks, information networks, telecommunications technology, data communications, protocols, data transfer, local area networks, Internet, data security, enterprises | Data communication & Networks |
| **26.** | Social media, marketing, media, marketing communication, Facebook, advertising, Internet, digital marketing, online communities, enterprises | Online advertising |
| **27.** | Databases, information management, data systems, computer programmes, programming, data, applications (computer programmes), SQL, database programs, final projects (education) | Information management |
| **28.** | Research, questionnaire survey, research methods, final projects (education), evaluation, development (active), use, services, objectives, enterprises | Research & Surveys |

**Supplementary Table 5**: 38 topic clusters and their manually assigned labels found in the job ads data set containing 107335 ads from ICT-field. For each cluster, we show 10 terms nearest to each cluster’s centroid. Table contains total 173 unique terms.

|  |  |  |
| --- | --- | --- |
| **No.** | **Cluster Terms** | **Cluster Label** |
| **1.** | Employment (legal relationship), working hours, employees, working life, recruitment of employees, staff, work, hiring, wages, enterprises | Employees & Staff |
| **2.** | Programming, computer programmes, software developers, programming languages, enterprises, information technology, planning and design, experiences (knowledge), development (active), know-how | Programming |
| **3.** | Information administration, municipalities, data systems, development (active), services, towns and cities, information technology, leadership (activity), tasks, organizations (systems) | Information administration |
| **4.** | Work communities, work, working life, work comfort, remote work, teamwork, success, expertise, development (active), business life | Working life |
| **5.** | Safety and security, security systems, data security, systems of supervision, safety technology, enterprises, security service sector, safety devices, passage control, data systems | Safety & Security |
| **6.** | Online services, Internet, services, enterprises, information technology, development (active), web pages, know-how, experiences (knowledge), information networks | Internet services & Information networks |
| **7.** | Information management, databases, data systems, information technology, management (control), data warehouses, development (active), enterprises, knowledge management, applications (computer programmes) | Information management |
| **8.** | Professional skills, know-how, education and training, staff, employees, working life, enterprises, experiences (knowledge), development (active), recruitment of employees | Professional skills |
| **9.** | Financial administration, electronic financial management, enterprises, development (active), personnel administration, business, data systems, services, customers, know-how | Financial administration |
| **10.** | Information technology sector, information technology, enterprises, services, know-how, recruitment of employees, experiences (knowledge), customers, customer service, employees | Information technology |
| **11.** | Access to employment, recruitment of employees, working life, employees, know-how, staff, information technology, enterprises, tasks, work | Employment |
| **12.** | Telecommunications technology, data communications, data communications networks, information networks, services, enterprises, mobile communication networks, information technology, information and communication technology sector, broadband networks | Data communication & Networks |
| **13.** | Business operations, business, enterprises, services, customers, development (active), marketing, know-how, corporate strategies, information technology | Business, Marketing & Corporate services |
| **14.** | Marketing, digital marketing, Internet, sale, social media, enterprises, electronic commerce, web pages, customers, communication | E-commerce & Digital marketing |
| **15.** | Infrastructures, consultancy agencies, management (control), language skills, languages, planning and design, projects, databases, programming, realization (active) | Consultancy |
| **16.** | Communication, customer experience, business life, customers, customer orientation, communicative skills, product development, leadership (activity), enterprises, innovations | Customer experience |
| **17.** | Personnel selection, recruitment of employees, jobseekers, access to employment, staff, labor (workforce), employees, working life, enterprises, employment (legal relationship) | Access to employment |
| **18.** | Business, development (active), enterprises, projects, know-how, services, leadership (activity), experiences (knowledge), data systems, customers | Business development |
| **19.** | Project management, project work, project leadership, projects, project planning, leadership (activity), planning and design, development (active), enterprises, customers | Project management |
| **20.** | Automation, installation, upkeep (servicing), service (maintenance), electricity sector, devices, education and training, enterprises, electrical engineering, real property | Automation |
| **21.** | Customer service, services, customers, information technology, enterprises, employees, recruitment of employees, experiences (knowledge), staff, know-how | Customer services |
| **22.** | Product development, innovations, technology, products, enterprises, planning and design, development (active), automation, know-how, experiences (knowledge) | Product development & design |
| **23.** | Cooperation (general), development (active), projects, business, enterprises, know-how, services, experiences (knowledge), leadership (activity), development projects | Business & Project cooperation |
| **24.** | Web pages, HTML, programming, JavaScript, Internet, programming languages, PHP, description languages, cascading style sheets, World Wide Web | Web programming |
| **25.** | Education and training, universities of applied sciences, universities, tasks, institutions of higher education, know-how, students, staff, development (active), teaching and instruction | Education & Learning |
| **26.** | Windows XP, Windows 95, Windows 7, Windows 98, Windows NT, Windows 8, information technology, operating systems, workstations (computing), enterprises | Windows OS |
| **27.** | Testing, testing methods, quality, computer programmes, software development, automation, quality assurance, quality control, software engineering, development (active) | Softwareengineering |
| **28.** | Public health service, health services, data systems, health sector, information administration, social welfare, development (active), social sector, hospital districts, federations of municipalities | Public health |
| **29.** | Managers and executives, leadership (activity), enterprises, experiences (knowledge), working life, know-how, staff, employees, recruitment of employees, development (active) | Enterprises& Leadership |
| **30.** | Tasks, presentations (introductions), central government, staff, defensive forces, employees, municipalities, legislation, know-how, wages | Legislation |
| **31.** | Success, organizations (systems), leadership (activity), self-evaluation, working life, development (active), enterprises, learning, career development, business life | Careerdevelopment |
| **32.** | Applications (computer programmes), design (artistic creation), planning and design, software design, product development, computer programmes, user-centeredness, management (control),customers, user interfaces | Softwaredesign |
| **33.** | Applications (documents), access to employment, workplaces, recruitment of employees, job application, employees, working life, staff, Internet, jobseekers | Recruitment |
| **34.** | Java ME, Java, Java EE, Java SE, programming, programming languages, Internet, enterprises, development (active), computer programmes | Javaprogramming |
| **35.** | Maintenance, servers, information networks, information technology, data systems, workstations (computing), Linux, operating systems, management (control), enterprises | Servers & OS |
| **36.** | Mobile devices, cell phones, operating systems, applications (computer programmes), software development, computer programmes, open-source code, software technology, Linux, product development | Mobile devices,Apps & OS |
| **37.** | Business, enterprises, development (active), services, projects, data systems, know-how, experiences (knowledge), leadership (activity), information technology | Business &Enterprises |
| **38.** | Customer orientation, customer service, customers, customership, marketing, services, sale, enterprises, customer relationship management, customer experience | Customership |