Samu Kumpulainen

Artificial General Intelligence - a systematic mapping study

Master's Thesis in Information Technology

October 7, 2020

University of Jyväskylä

Faculty of Information Technology

Author: Samu Kumpulainen

Contact information: samu.p.kumpulainen@student.jyu.fi

Supervisor: Vagan Terziyan

Title: Artificial General Intelligence - a systematic mapping study

Työn nimi: Yleistekoäly - systemaattinen kirjallisuuskartoitus

Project: Master's Thesis

Study line: Mathematical Information Technology

Page count: 22+0

Abstract: In this thesis, a systematic mapping study is performed on the field of artificial general intelligence. The goal of the study is to gain insight about the recent developments in the study field. This includes the focus points of the current reserch, possible research gaps, and how the research itself is conducted. TODO: more accurate, proper abstract

Keywords: Master's Theses, AGI, AI, artificial intelligence, systematic literature mapping, mapping study

Suomenkielinen tiivistelmä:

Tässä suunnitelmassa käydään läpi pro gradu -tutkielman mahdollista aihetta ja tutkimustapaa. TODO: translate when abstract done

Avainsanat: Pro Gradu,tutkielma, AI, tekoäly

List of Figures

Contents

1 Introduction

The thesis will be a systematic research mapping on the field of Artificial General Intelligence (AGI). The goal of the thesis is to identify the themes and subfields of AGI research in recent years, what is being researched recently, and what kind of gaps exist on the field. For a while the AGI field was not so active and the more specific approaches, 'narrow AI', grew in popularity. Recently, however, the wider, more general artificial intelligence has been regaining interest. This kind of mapping study would be needed as the research field is complex and there is no clear presentation of the current trends and focal points. Creating this king of overview would be a valuable asset for future research, as it would enable focusing the research on areas less ventured. Furthermore, if an interesting subtopic comes up during the process of mapping, more focus may be directed towards that in form of more traditional systematic literature review. This option is left for further consideration.

What each chapter is about, how the thesis is structured etc.

2 Artificial General Intelligence

TODO: Here some chapter introducing text

2.1 History of Artificial Intelligence

Even though the idea of autonomous machinery has been around since the ancient Greek (), AI's origins are set around in the 1940s. At the time, American science fiction author Isaac Asimov wrote numerous novels and short stories about conscious robots and technology's relation to humankind. His work has inspired countless people in the field's of AI and computer science since then (Haenlein and Kaplan July 2019). Also in the 1940s, mathematician Alan Turing's work on Britain's code breaking efforts lead to the creation of first electromechanical computer, The Bombe (Haenlein and Kaplan July 2019). Turing later gave lectures and wrote an article titled *"Computing Machinery and Intelligence"* (1950), in which he presented several ideas later prevalent in AI field, including the "Imitation game", a test to measure the intelligence of a machine (Russell and Norvig 2009). This later became well known as the Turing test.

The term Artificial Intelligence was coined in 1956 during a two-month workshop *Darthmouth Summer Research Project on Artificial Intelligence*, organized by John McCarthy and Marvin Minsky (Haenlein and Kaplan July 2019). The participants of the workshop would later become the most prominent figures of AI research. During DSRPAI two researchers, Allen Newell and Herbert Simon presented Logic Theorist, their existing reasoning program, capable of proving multiple mathematical theorems (Russell and Norvig 2009). Based on this work the two later created General Problem Solver, GPS, which could solve simple puzzles like Towers of Hanoi using human like recursive approach (Newell, Shaw, and Simon 1959). The early days of AI research produced many similar results in different areas. IBM's Arthur Samuel created AI programs that learned to play checkers at a strong amateur level (Russell and Norvig 2009). John McCarthy's 1958 paper titled "Programs with common sense", describes Advice Taker, a complete but hypothetical AI system with general knowledge about the world and deductive processes to manipulate it. The paper is still thought to be relevant today. McCarthy's system was able to acquire new skills in previously unknown areas without being reprogrammed.

During these years also work on the neural networks started to gain interest. The initial work of McCulloch and Pitts (1943), later demonstrated by Hebb (Russell and Norvig 2009), showed that a neural network is capable of learning. In 1960s Rosenblatt's work on perceptrons and Widrow and Hoff's LMS algorithm were some of the biggest advances in the area (Widrow and Lehr 1995). The next great discovery that would propel the neural networks into the focal point of AI research would happen in the mid-1980s when the backpropagation algorithm originally presented by Bryson and Ho in 1969 was rediscovered by multiple independent groups (Russell and Norvig 2009). Backpropagation is one of the most widely used algorithms for training neural networks these days for its relative power and simplicity (Rumelhart et al. 1995).

History of artificial intelligence contains occasional periods of reduced interest and funding. These so called "AI winters" are a result of high expectations collapsing under criticism. First period that can be considered an AI winter started in the 1970s, and Russell and Norvig (2009) present the following possible reasons for it: Firstly, the early programs knew nothing about their context, and solved the problems via syntactic manipulations. This was especially apparent on machine translation projects. As a language cannot be fully understood without knowing the full context of the sentences and other nuances of the language, accurate translation proved to be a difficult task. Failed translation efforts lead to funding cuts in the US. Second difficulty pointed out by (Russell and Norvig 2009) was the sheer complexity of the target problems. As the early AI programs were focused on simple tasks, finding a solution by trial and error was possible in practice. But as the problems became more complex, "combinatorial explosion" issue became more apparent. The issue was also discussed in British scientist James Lighthill's report on the state of the AI (1973). The report is considered to be one of the main reasons why the British government decided to cut all AI funding in all but two universities. Lastly, the limitations of the data structures used in AI field, such as perceptrons, restricted the capabilities of the solutions. According to Russell and Norvig (2009) this lead to funding cuts also in the neural network research.

During and after the first AI winter, there was a considerable amount of research relating to expert systems (Russell and Norvig 2009). These systems perform their tasks in a way similar to human experts in the specific, narrow domain, relying on a knowledge encoded into a set of rules (Myers 1986). This style of AI research was inspired by the success of DENDRAL (Buchanan, Feigenbaum, and Lederberg 1968), a system developed at Stanford by Ed Feigenbaum, Bruce Buchanan and Joshua Lederberg. DENDRAL's purpose was to use data from mass spectrometer to infer the structure of a given molecule. MYCIN, developed in the 1970s (Shortliffe et al. 1975), incorporated domain knowledge acquired through expert interviews, with the uncertainty of medical evaluation taken into account via certainty factors (Russell and Norvig 2009).

Expert systems gained commercial interest, leading to increased research and adoption in the industry. Government investments in Japan lead to increased funding in United States and Britain, leading to an AI boom in the 1980s (Russell and Norvig 2009). After the boom, at the end of the 1980s, the second AI winter arrived. Participation in AI conferences dropped, several of the new AI companies met their end, as did the AI research divisions in larger hardware and software companies (Nilsson 2009). The imminent burst of the bubble was foreseen by several leading researchers, but their warnings didn't have considerable effect (Nilsson 2009).

According to Russell and Norvig (2009), around this time the AI field started to adopt the scientific method. This means the earlier ways of proposing completely new theories based on vague evidence or oversimplified examples have been replaced by basis on existing theories, repeatable experiments, and real-world examples. This newly discovered open-mindedness then lead to a complete new ways of looking at the AI research. AI solutions based on existing theories, such as speech recognition based on hidden Markov models, enables the researchers to build on the rigorous mathematical theory behind it (Russell and Norvig 2009). Work of Judea Pearl (1988) and Peter Cheeceman (1985) on the probabilistic reasoning lead to it being accepted back into the AI field. Later Pearl's Bayesian networks have been used to handle uncertainty in AI problems. They are graphical models that join probabilistic information and dependencies to events, enabling inference using probabilistic methods (Goertzel and Pennachin 2007).

In the 21st century artificial intelligence research has been steadily growing. According to

(Liu et al. 2018), not only has the amount of publications in the field been increasing, but also the collaboration between researchers. The study also deduces that that AI has become more open-minded and popular, as the rate of self-references is reducing. One reason for the rising popularity on the field is the success that narrow AI solutions have presented in multitude of problems. For example, in classical game of Go, program called AlphaGo developed by Google-owned DeepMind, defeated the world champion Lee Sedol in 2015 (Silver et al. 2016). Due to Go's computationally complex nature this was a impressive feat previously thought impossible. Later DeepMind developed even more advanced versions of AlphaGo, called AlphaGo Zero, and generalized AlphaZero, which could even play Shogi and Chess on superhuman level (Silver et al. 2018).

Recent years majority of the field has been focusing on the narrow AI approaches (Goertzel and Pennachin 2007). However, the interest in the classical, strong AI has also been increasing. This can be seen in the publications from many influential AI researchers. Authors like John McCarthy (2007), Nils Nilsson (2005) and Marvin Minsky (2007) have voiced their opinions that efforts to create a more general AI should be pursued. There are several terms used regarding these efforts. **Human-level Artificial Intelligence** (HLAI) aims to reach "human-level intelligence" and common sense, a goal that according to Marvin Minsky (2004) can be reached by not using any single method, but a combination of different resources and methods. Similar term is **Artificial General Intelligence** (AGI) presented by Ben Goertzel and Casio Pennachin (2007). The goal of AGI is similar to HLAI, to create an AI system that can express general intelligence instead of being locked into a single domain. On the next chapter this general approach is presented in more detail, as it is the focus of this thesis.

2.2 Definition

In order to be able to define AGI, or artificial intelligence in general, one must first consider the definitions of intelligence in general. The exists many different definitions, in many different branches of science. Legg and Hutter (2007) list over 60 definitions collected from various academic sources. These include, for example, *"the general mental ability involved in calculating, reasoning, perceiving relationships and analogies, learning quickly, storing* *and retrieving information,using language fluently, classifying, generalizing, and adjusting to new situations."* (Columbia Encyclopedia, sixth edition, 2006), *"that facet of mind underlying our capacity to think, to solve novel problems, to reason and to have knowledge of the world"* (Anderson 2006), and *"Intelligence is the ability for an information processing system to adapt to its environment with insufficient knowledge and resources."* (Wang 1995). Based on the aforementioned collection of definitions, Legg and Hutter (2007) have formed the following definition: *"Intelligence measures an agent's ability to achieve goals in a wide range of environments"*. This gives us a single definition which encompasses the common traits in intelligence definitions.

Artificial General Intelligence, sometimes referred as "strong ai", according to Goertzel and Pennachin (2007) means *"AI systems that possess a reasonable degree of self-understanding and autonomous self-control, and have the ability to solve a variety of complex problems in a variety of contexts, and to learn to solve new problems that they didn't know about at the time of their creation."*. It can be seen that an agent fulfilling this definition also possesses the intelligence defined in the previous chapter. In this thesis terms artificial general intelligence and human-level artificial intelligence are treated as synonyms, as they pursue more or less the same goal of general intelligence. Goertzel and Pennachin (2007) suggest that the term AGI is more fitting to the area than HLAI as it human-like approaches are not necessarily used.

The reason general intelligence is specified instead of plain intelligence is that there is a need to differentiate it from the domain specific artificial intelligence, also known as "narrow AI" or "weak AI", that has become prevalent in AI research in recent past. Terms strong AI and weak AI were coined by John Searle in 1980 (John et al. 1980). Narrow AI means smart solutions that may learn and improve their performance through training, but they are only focused on specific type of problems in a specific context. Examples of such AI include chess engines, autonomous vehicles, and natural language processing. These solutions may outperform human capabilities, but only in their limited tasks. When presented a problem outside their domain, they usually perform poorly. As the above definition by Goertzel and Pennachin describes, AGI is able to function on different context and tasks without separate human intervention and reconfiguration.

As The AGI community is diverse and there are multitude of opinions on the best approaches and the goals that should be pursued in the research, several possible roadmaps have been presented in an attempt to create a common basis for the discussion and research of human-level artificial general intelligence. In (Adams et al. March 2012) a high level roadmap with AGI's initial required capabilities and scenario-based milestones is suggested, building on previous work and workshops organized in 2008 and 2009. Presented scenarios can be used to measure the progress and capabilities of AGI restricting the progress of different approaches to a single test situation (Adams et al. March 2012). More concrete example is provided by Ben Goertzel and Gino Yu, who outline creation of a AGI-oriented cognitive architecture based on existing CogPrime architecture (Goertzel and Yu 2014). A simultaneous development of multiple AGI-style applications is suggested to maintain the generality of intelligence. CogPrime is implemented with OpenCog framework, developed by OpenCog Foundation and AI researcher Ben Goertzel. OpenCog is an attempt to create an open source framework for artificial general intelligence (*The Open Cognition Project;* Goertzel 2012).

TODO: Add something like: One motivation behind this thesis is to find out if these roadmaps and "common ground" have actually lead to anything concrete, or has their effort been for nothing.

3 Systematic literature mapping process

Systematic literature mapping is a secondary study method that helps to identify the focal points and research gaps in the subject area, providing an overview of previous research (Petersen et al. 2008). This chapter introduces the mapping method, describing each phase of the research process. The key differences with a more popular study method, systematic literature review (SLR) are presented.

3.1 Research method description

Systematic mapping is a common research method used in fields such as evidence based medicine, but have until recently been rare in software engineering (Petersen et al. 2008). According to Kitchenham et al. (Kitchenham et al. 2010),

The systematic literature mapping in this thesis is following the method guidelines presented by Petersen et al (2008), later updated in by Petersen, Vakkalanka and Kuzniarz (2015). The mapping process overview can be seen in figure 1. It consists of five separate phases: **definition of research questions, conducting search, screening of papers, keywording, and data extraction and mapping.** Each phase produces a subresult to be used in the next one. This process results in a systematic map of the area. This can and should be further visualized using for example bubble graphs (Petersen et al. 2008). Visualization enables easier recognition of research gaps and focus points in the target area.

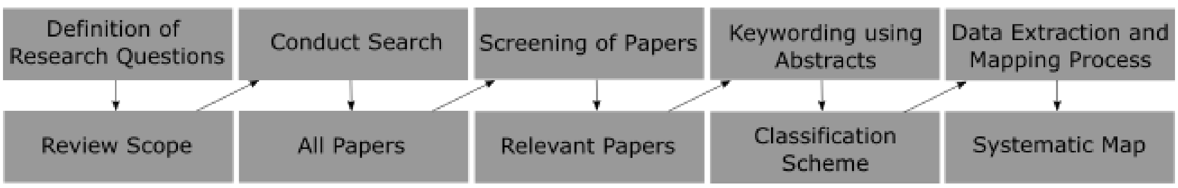

Figure 1. Process model (Petersen et al. 2008)

The following text is very raw, will be edited properly later

The process begins by defining focused research questions that are aligned with the goal of the study. The goal is usually to create a general overview of the research area, and to identify the type and quantity of research. Unlike in more focused systematic literature reviews, the research questions of mapping studies are less focused and cover a broader scope.

The next phase is the initial material search, which can be conducted multiple ways. Search strings can be formulated from the research questions, and used on academic databases and search engines. For example, databases such as IEEE Explore and ACM, as well as aggregators like Google Scholar. As the goal is to achieve broad overview of the research, study outcomes are not taken into account. The search can also be conducted manually on specific journals and conference publications that cover the target area. This is the approach used in this thesis. This search method also facilitates the initial exclusion of irrelevant papers, limiting the possible material to the similar fields.

After the initial material has been gathered, it is further refined by excluding papers not relevant to answering the research questions. Separate criteria for both inclusion and exclusion is used to find the papers fit for further analysis. Criteria can be meta-level, such as publication source, year and format, and content-based, considering the article itself.

Once the final set of papers is narrowed down and determined, keywording is performed. First, the research papers' abstracts are analyzed by looking up possible keywords and common concepts from them. The keywords of each paper are then combined together. If abstracts' are low quality, the introduction and conclusion of the paper can also be examined. On some cases inspecting the whole article might be required. Different facets can be used, for example topical and one reflecting the research approach (topic-independent). After the final set of keywords is chosen, they are clustered into categories representing the population.

In the final phase of the mapping process, the papers are sorted into the classification scheme. The schema may evolve during the data extraction process, changing the categories to match the article population. Once the categorization is complete, the resulting frequencies of papers can be presented either via visualization or summary statistics. Visualization using e.g. bubble plots is preferred, as it is a powerful way to represent the information and map of the field.

3.2 Difference with other secondary studies

3.3 Mapping studies in field of IT

4 Conducting the literature mapping

- To ensure study is conducted with rigor, should it be documented and reported very carefully.

4.1 Background/reasoning behind method choice

4.2 Research questions

The following research questions...

1. How much, and what type of research is done in the field of AGI?
2. Where is the AGI research focused on?
3. Has there been any major breakthroughs?
4. Where and when were the studies published?

4.3 Sources and databases used

- Journal listing and their date ranges etc.

- Why journals were chosen, information about them, reliability, area coverage, jufo rankings etc.

- search terms here or another section?

- table showing used search phrases?

4.4 conducting search

Search phrases are used on different databases, limiting the papers to amount possible to handle

Search can be also done by inspecting relevant journals (like here)

- evaluation of search: test set of known papers, possible here?

4.5 Criteria for inclusion, exclusion

Criteria is presented, and papers are further narrowed down.

4.6 Keywording

Papers are further analyzed, keywords are extracted from abstracts and text.

4.7 Data extraction and mapping

keywords are used to create a mapping, working iteratively. Papers are categorized based on the emergent mapping.

4.8 Source material control

- How the papers were handled - How graphs etc. were made - Other meta level information

5 Results and analysis

The results of the thesis will be a clear overview of the recent research in the field of artificial general intelligence in form of classification data, visual graphs and further synthesis. As mentioned earlier, if an interesting topic presents itself during the process of mapping, it may be further examined in a more focused way, such as literature review. This will be considered at that time.

5.1 Validity threats?

- performed alone, researcher bias etc (multiple people not possible here), how rigor is achieved, repeatability,

Good source on reliability Wohlin et al. 2013

5.2 Results of literature mapping

- Graphs and other visualization, bubble graphs are useful. - Key topics should be described shortly?

5.3 Possible continuation research

- List of most prominent topics for further research

6 Conclusion

In this thesis, a systematic literature mapping was conducted on the field of artificial general intelligence. Results of the study showed that ....

Bibliography

Adams, Sam, Itmar Arel, Joscha Bach, Robert Coop, Rod Furlan, Ben Goertzel, J. Storrs Hall, et al. March 2012. ”Mapping the Landscape of Human-Level Artificial General Intelligence”. *AI Magazine* 33, number 1 (): 25–42. https://doi.org/10.1609/aimag.v33i1.2322. https://www.aaai.org/ojs/index.php/aimagazine/article/view/2322.

Anderson, M. 2006. ”Intelligence”. *MS Encarta online encyclopedia.*

Buchanan, B, E Feigenbaum, and J Lederberg. 1968. ”Heuristic DENDRAL-A program for generating explanatory hypotheses in organic chemistry.”

Cheeseman, Peter C. 1985. ”In Defense of Probability.” In *IJCAI,* 2:1002–1009.

Goertzel, B. 2012. ”CogPrime : An Integrative Architecture for Embodied Artificial General Intelligence”.

Goertzel, Ben, and Cassio Pennachin. 2007. *Artificial general intelligence.* Volume 2. Springer.

Goertzel, Ben, and Gino Yu. 2014. ”From here to AGI: A roadmap to the realization of human-level artificial general intelligence”. *2014 International Joint Conference on Neural Networks (IJCNN),* 1525–1533.

Haenlein, Michael, and Andreas Kaplan. July 2019. ”A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence:” [**inlang**en]. *California Management Review* (). https://doi.org/10.1177/0008125619864925. https://journals-sagepub-com.ezproxy.jyu.fi/doi/full/10.1177/0008125619864925?casa_token=cJ1YEP0BQQQAAAAA%3ANQqrLNDBpfNEU7WmxeiyQ178cT8ONGu70KePNj4i-XuyIxQTKlbnFz11d1rXHTtf9kNb_GKVXmnZrw.

John, Searle, et al. 1980. ”Minds, Brains and Programs”. *Behavioral and Brain Science* 3:417–424.

Kitchenham, Barbara, Rialette Pretorius, David Budgen, O Pearl Brereton, Mark Turner, Mahmood Niazi, and Stephen Linkman. 2010. ”Systematic literature reviews in software engineering–a tertiary study”. *Information and software technology* 52 (8): 792–805.

Legg, Shane, Marcus Hutter, et al. 2007. ”A collection of definitions of intelligence”. *Frontiers in Artificial Intelligence and applications* 157:17.

Lighthill, James. 1973. ”Artificial intelligence: A general survey”. In *Artificial Intelligence: a paper symposium,* 1–21. Science Research Council London.

Liu, Jiaying, Xiangjie Kong, Feng Xia, Xiaomei Bai, Lei Wang, Qing Qing, and Ivan Lee. 2018. ”Artificial intelligence in the 21st century”. *IEEE Access* 6:34403–34421.

McCarthy, John. 2007. ”From here to human-level AI”. *Artificial Intelligence* 171 (18): 1174–1182.

McCulloch, Warren S, and Walter Pitts. 1943. ”A logical calculus of the ideas immanent in nervous activity”. *The bulletin of mathematical biophysics* 5 (4): 115–133.

Minsky, Marvin. 2007. *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind.* Simon / Schuster.

Minsky, Marvin L, Push Singh, and Aaron Sloman. 2004. ”The St. Thomas common sense symposium: designing architectures for human-level intelligence”. *Ai Magazine* 25 (2): 113–113.

Myers, Ware. 1986. ”Introduction to expert systems”. *IEEE Computer Architecture Letters* 1 (01): 100–109.

Newell, Allen, John C Shaw, and Herbert A Simon. 1959. ”Report on a general problem solving program”. In *IFIP congress,* 256:64. Pittsburgh, PA.

Nilsson, Nils J. 2005. ”Human-level artificial intelligence? Be serious!” *AI magazine* 26 (4): 68–68.

———. 2009. *The quest for artificial intelligence.* Cambridge University Press.

Pearl, Judea. 1988. ”Probabilistic reasoning in intelligent systems: Networks of plausible inference”. *Morgan Kauffmann Publishers Inc.*

Petersen, Kai, Robert Feldt, Shahid Mujtaba, and Michael Mattsson. 2008. ”Systematic mapping studies in software engineering”. In *12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12,* 1–10.

Petersen, Kai, Sairam Vakkalanka, and Ludwik Kuzniarz. 2015. ”Guidelines for conducting systematic mapping studies in software engineering: An update”. *Information and Software Technology* 64:1–18.

Rumelhart, David E, Richard Durbin, Richard Golden, and Yves Chauvin. 1995. ”Backpropagation: The basic theory”. *Backpropagation: Theory, architectures and applications,* 1–34.

Russell, Stuart, and Peter Norvig. 2009. *Artificial Intelligence: A Modern Approach.* 3rd. Prentice Hall Press. ISBN: 0136042597.

Shortliffe, Edward H, Randall Davis, Stanton G Axline, Bruce G Buchanan, C Cordell Green, and Stanley N Cohen. 1975. ”Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system”. *Computers and biomedical research* 8 (4): 303–320.

Silver, David, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. ”Mastering the game of Go with deep neural networks and tree search”. *nature* 529 (7587): 484–489.

Silver, David, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. 2018. ”A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play”. *Science* 362 (6419): 1140–1144.

The Open Cognition Project. Accessed: 2020-09-24. https://wiki.opencog.org/w/The_Open_Cognition_Project.

TURING, A.M. 1950. ”Computing machinery and intelligence”. *Mind* 59 (236): 433–460.

Wang, Pei. 1995. ”On the working definition of intelligence”. *Center for Research on Concepts and Cognition CRCC, Indiana University.*

Widrow, Bernard, and Michael A Lehr. 1995. ”Perceptrons, Adalines, and backpropagation”. *Arbib* 4:719–724.

Wohlin, Claes, Per Runeson, Paulo Anselmo da Mota Silveira Neto, Emelie Engström, Ivan do Carmo Machado, and Eduardo Santana de Almeida. 2013. ”On the reliability of mapping studies in software engineering”. *Journal of Systems and Software* 86 (10): 2594–2610. ISSN: 0164-1212. https://doi.org/https://doi.org/10.1016/j.jss.2013.04.076. http://www.sciencedirect.com/science/article/pii/S0164121213001234.