

June 2006

Index

Introduction

Quantifying
knowledge

The *h* index

Pros & cons

Practical *h*

Math facts

Usage

The *h* index, the quest for an optimum metric

<http://neo.lcc.uma.es>



One of the research lines of the GISUM group

Enrique Alba
University of Málaga, Spain



Index

➔ Introduction

Quantifying
knowledge

The *h* index

Pros & cons

Practical *h*

Math facts

Usage

Introduction

Optimize

- **The search for an objective metric to quantify the cumulative impact and relevance of an individual's scientific research output**

Subject to

- **Should be easy to understand and measure**
- **Should be applicable to different domains in science**
- **Should measure the impact of the research**
- **Should avoid inflating distortion by 'big hits'**
- **Should be applicable to scientists of different ages**
- **Should be useful for evaluation and comparison (recruitment, salary complements, grants, awards, etc.)**



Index

Introduction

→ Quantifying knowledge

The h index

Pros & cons

Practical h

Math facts

Usage

Quantifying the Knowledge

General thoughts

- The publication data of an individual is the most usual source of information
- But it is not the only one: the citation record of a researcher is also very important

Thus we can quantify relevant information

- N_p : Number of papers published over n years
- N_c^j : Number of citations for each paper j
- Journals where the papers were published
- Impact factor of a journal (ISI or others)
- Others (trend of a journal, domain rankings, etc.)



Index

Introduction

Quantifying
knowledge

→ The h index

Pros & cons

Practical h

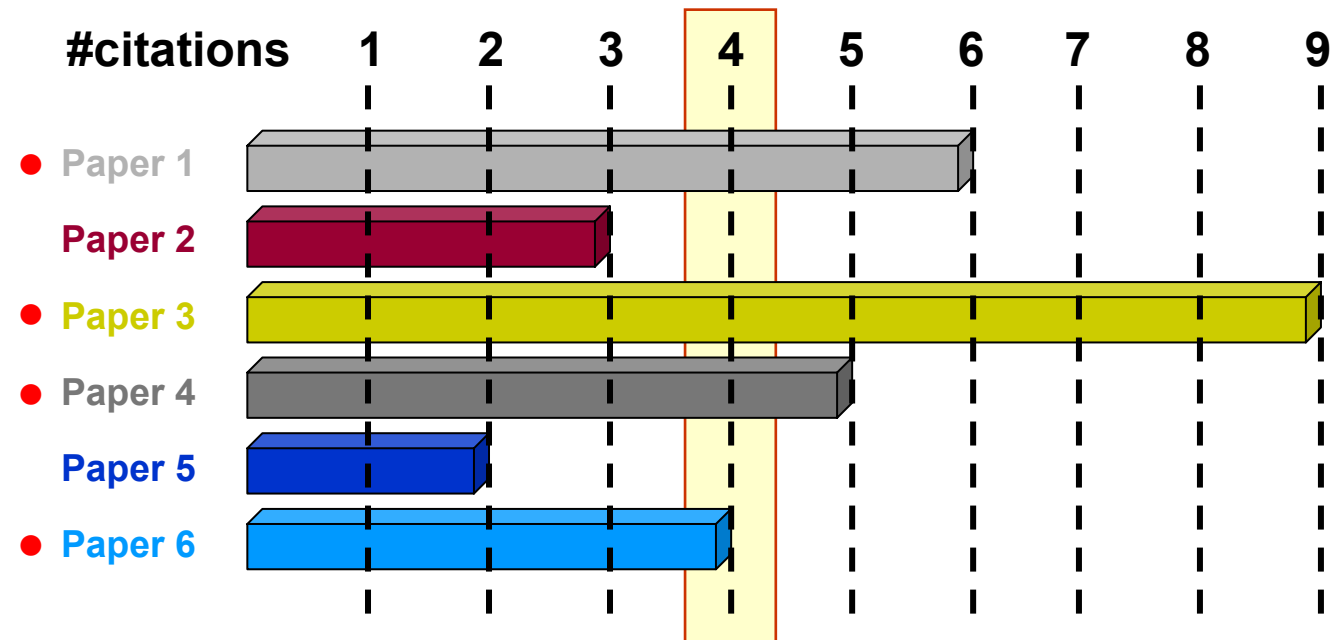
Math facts

Usage

The h index

Definition of the h index

A scientist has index h if h of his/her N_p papers have at least h citations each, and the rest of $(N_p - h)$ papers have no more than h citations each





Index

Introduction

Quantifying
knowledge

The h index

➔ Pros & cons

Practical h

Math facts

Usage

Advantages and Disadvantages of the h index

Why h is preferable to other single-number criteria:

1. Total number of papers (N_p)

PROs: measures productivity

CONs: does not measure importance nor impact of papers

2. Total number of citations ($N_{c,tot}$)

PROs: measures total impact

CONs: hard to find, inflated by 'big hits' or by many coauthors

3. Average citations per paper ($N_{c,tot} / N_p$)

PROs: measures the average impact of publications

CONs: hard to find, rewards/penalize low/high productivity

4. Number of 'significant' papers ($> y$ citations)

PROs: no problems like before, it gives an idea of broad impact

CONs: favor/disfavor people arbitrarily, needs tuning for ages

5. Number of citations to each of the q most cited

PROs: overcomes many of the precedent disadvantages

CONs: favor/disfavor people arbitrarily, not a single number



Index

Introduction

Quantifying
knowledge

The h index

Pros & cons

➔ Practical h

Math facts

Usage

Practical Issues

Advantages of h -index:

1. Measures the broad impact of an individual's work
2. Avoids most of the mentioned disadvantages (not all of them)
3. Easy to find: rank papers by 'times cited' in Thomson ISI (only this?)
4. Gives a rough estimate of the total number of citations

Practical application:

- A. Two individuals with similar h are comparable in terms of their overall scientific impact, even if their total number of papers and citations are very different
- B. For two individuals (of the same scientific age) with similar number of total papers and of total citation count but with very different h values, the one with the higher h is likely to be the more accomplished scientist



Index

Introduction

Quantifying
knowledge

The h index

Pros & cons

Practical h

➔ Math facts

Usage

Mathematical Facts

Direct Implications (I)

1. The total number of citations is $N_{c,tot} = a \cdot h^2$

- The lower bound is h^2 , ignoring #citations: above/below h
 $a \sim [(1+c/p)^2] / [2c/p]$, p papers earning c citations (per year)
- Parameter a ranges between 3 and 5 (in physics)

2. Assuming linear grow with time: $h \sim m \cdot n$

- For a not too small number of years n , $h = [c / (1+c/p)] \cdot n$
- Slope (m) is useful to compare scientists of different seniority

3. The minimum value for a is 2 (when $c=p$)

- Papers with more/less of h citations contribute eq. to $N_{c,tot}$
- The value of a will be larger for both $c > p$ and for $c < p$

4. Most (all) contributions come from highly cited papers when $c > p$

- Note: for $c < p$ the $(N_p - h)$ papers are giving the largest contribution to $N_{c,tot}$



Index

Introduction

Quantifying knowledge

The h index

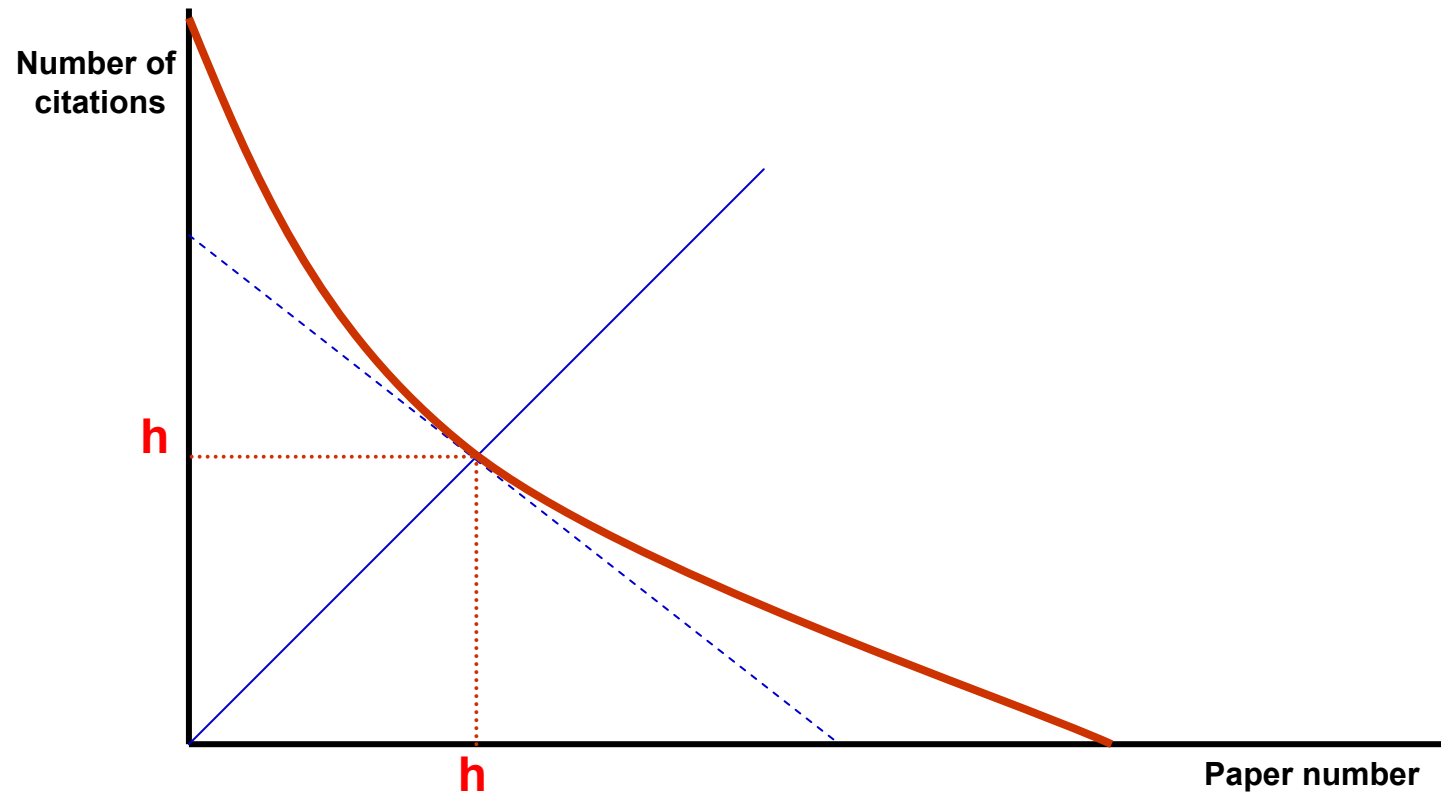
Pros & cons

Practical h

➔ Math facts

Usage

Mathematical Facts



The intersection of the 45 degree line with the curve giving the number of citations versus the paper number gives h . The total number of citations is the area under the curve. Assuming the second derivative is non-negative everywhere, the minimum area is given by the distribution indicated by the dotted line, yielding $a=2$ (being the slope = -1)



Index

Introduction

Quantifying
knowledge

The h index

Pros & cons

Practical h

➔ Math facts

Usage

Mathematical Facts

Direct Implications (II)

5. In case a researcher stops publishing

- The linear model makes h to grow constantly during h/c years and then stays constant, since now all published papers contr.
- In a more realistic model h will level off smoothly

6. Not all papers contribute to h

- The ones with fewer than h citations won't contribute, like those written late in a career or in not well known media

7. Papers earn citations over a limited period

- Papers contributing to a career are different in diff. moments

8. The h index cannot decrease with time

- Papers with h citations are in risk of not counting more
- 'Sleeping beauties' are papers that drop and come back again



Index

Introduction

Quantifying
knowledge

The h index

Pros & cons

Practical h

Math facts

➔ Usage

General Usage

Practical examples ($m = \frac{c}{1 + c/p}$)

1. $m \sim 1$

i.e. $h=20$ after $n=20$ years characterizes a successful scientist

2. $m \sim 2$

i.e. $h=40$ after $n=20$ years characterizes outstanding scientists,
likely to be found only at top universities or research labs

3. $m \sim 3$ or higher

i.e. $h=60$ after $n=20$ years, or $h=90$ after $n=30$ years characterizes
truly unique individuals

The m parameter ceases its usefulness if productivity severely drops, but the h -parameter is cumulative and may continue to increase even after a stop



Index

Introduction

Quantifying knowledge

The h index

Pros & cons

Practical h

Math facts

➔ Usage

Examples in Physics

Suggestions in physics (assuming $m \sim 1$):

1. $h \sim 10..12$

Might be typical values for tenure (associate professor)

2. $h \sim 15..20$

Might be used to fellowship in the American Physical Society

3. $h \sim 18$

Might be used to promote to full professor

4. $h \sim 45$

Might be used to enter the US National Academy of Sciences

There can always be exceptions to rules, especially in life-changing decisions such as granting or denying of tenure (or civil servants in research)



Index

Introduction

Quantifying knowledge

The h index

Pros & cons

Practical h

Math facts

➔ Usage

Fields, Self-citations and Groups

Typical h -values will be different in different fields:

- Average number of references in a paper
- Average number of papers per scientist in the field
- Number of scientists in the field

A high h means high accomplishment, but the converse is not necessarily always true

- What if a few seminal papers with very large number of citations
- What if papers with many coauthors provoke a large h index

Self citations should be eliminated

- Their effect in the h index is small (irrelevant for papers below h)
- To correct, a paper with $h+n$ citations and more than n self citations should be dropped from the h count ($h--$)

The h index of a group is larger than individual's ones but smaller than the sum of individual's h indices



Index

Introduction

Quantifying
knowledge

The h index

Pros & cons

Practical h

Math facts

➔ Usage

Hands on the h Index

The are are many sources for consulting citations:

- ISI Web of knowledge: <http://portal.isiknowledge.com/>
- SCOPUS citation tracker: <http://www.scopus.com/>
- Citeseer archives: <http://citeseer.ist.psu.edu/>
- IEEEXplore crossref search (Google): <http://ieeexplore.ieee.org>
- Science direct (SCOPUS): <http://www.sciencedirect.com/>
- Google scholar: <http://scholar.google.com/>
- DBLP CS biblio. : <http://www.informatik.uni-trier.de/~ley/db/>
- ACM digital library citings (Google): <http://portal.acm.org>

Search tips

- Only ISI and SCOPUS provide some tools for computing the h index
- The rest of sources must be seek paper by paper, tracking the cites (!)
- No automatic tool exists to exclude self references (filtered at hand)
- An automatic bibref tool should help, but just a few people do this.
- Search in Google for the exact article names to find new references



Index

Introduction

Quantifying
knowledge

The *h* index

Pros & cons

Practical *h*

Math facts

➔ Usage

Hands on the *h* Index: ISI

Using ISI (I)

- Go to ISI Web of knowledge and select “Web of Science”
- Select “Cited Ref Search”
- Enter your name in the search box (e.g. alba e*)
- Select your articles page by page and then click on “Finish Search>>”

ISI Web of KnowledgeSM Web of Science

FINISH SEARCH The completed search will be added to the search history.

(Limit by language and document type)

CITED REFERENCE INDEX

Go to Page: 1 of 4

References 1 -- 50

SELECT PAGE SELECT ALL* CLEAR ALL or select specific references from the list.

When desired references have been selected from all pages, click FINISH SEARCH to complete your search.

Select	Times Cited**	Cited Author	Cited Work [SHOW EXPANDED TITLES]	Year	Volume	Page	Article ID	View Record
<input type="checkbox"/>	1	ALBA E	2 S GEOL REG ARG AC	1979	1	349		
<input type="checkbox"/>	1	ALBA E	3RD P INT C INF OBST	1988				
<input type="checkbox"/>	1	ALBA E	5121629	1992	US			View Record
<input type="checkbox"/>	2	ALBA E	7 C SOC ESP ONC MED	1999				
<input type="checkbox"/>	1	ALBA E	ACT 12 JORN PAR	2001		105		
<input checked="" type="checkbox"/>	1	ALBA E	ADV SOFT COMPUTING H	2002	113	211		
<input type="checkbox"/>	1	ALBA E	ALPHABETICAL INDEX O	1950				
<input type="checkbox"/>	1	ALBA E	ALYTES	1983	1	225		
<input type="checkbox"/>	2	ALBA E	AM J CLIN ONCOL-CANC	1996	19	140		View Record
<input type="checkbox"/>	1	...Alba E	AM J CLIN PATHOL	2006	125	111		View Record
<input type="checkbox"/>	1	ALBA E	AM J EMERG MED	1992	10	55		
<input type="checkbox"/>	14	ALBA E	AM J EMERG MED	1992	10	64		View Record
<input type="checkbox"/>	2	...Alba E	AM J ROENTGENOL	2005	184	812		View Record
<input type="checkbox"/>	9	...Alba E	ANN ONCOL	2004	15	79		View Record
<input type="checkbox"/>	3	...Alba E	ANN ONCOL	2004	15	1798		View Record
<input type="checkbox"/>	13	...Alba E	ANN ONCOL	2003	14	867		View Record
<input type="checkbox"/>	14	...Alba E	ANN ONCOL	1999	10	289		View Record
<input type="checkbox"/>	7	ALBA E	ANN ONCOL	1992	3	31		View Record



Index

Introduction

Quantifying knowledge

The h index

Pros & cons

Practical h

Math facts

➔ Usage

Hands on the h Index: ISI

Using ISI (II)

- Select all your the articles (at a time or filter to ensure that it's you)
- On the right hand menu select “Sort by: Times Cited”... and good luck!

ISI Web of KnowledgeSM Web of Science Signed In

23 results found Go to Page: 1 of 3
 Records 1 -- 10 Show 10 per page

Use the checkboxes to select records for output. See the sidebar for options.

<input checked="" type="checkbox"/>	1.	Alba E, Tomassini M Parallelism and evolutionary algorithms IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION 6 (5): 443-462 OCT 2002 Times Cited: 35
<input checked="" type="checkbox"/>	2.	Alba E, Troya JM Analyzing synchronous and asynchronous parallel distributed genetic algorithms FUTURE GENERATION COMPUTER SYSTEMS 17 (4): 451-465 JAN 2001 Times Cited: 2
<input checked="" type="checkbox"/>	3.	Alba E, Troya JM Influence of the migration policy in parallel distributed GAs with structured and panmictic populations APPLIED INTELLIGENCE 12 (3): 163-181 MAY 2000 Times Cited: 6
<input checked="" type="checkbox"/>	4.	Alba E, Troya JM Improving flexibility and efficiency by adding parallelism to genetic algorithms STATISTICS AND COMPUTING 12 (2): 91-114 APR 2002 Times Cited: 5
<input checked="" type="checkbox"/>	5.	Alba E Parallel evolutionary algorithms can achieve super-linear performance INFORMATION PROCESSING LETTERS 82 (1): 7-13 APR 15 2002 Times Cited: 5
<input type="checkbox"/>	6.	Alba E, Dorransoro B The exploration/exploitation tradeoff in dynamic cellular genetic algorithms IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION 9 (2): 126-142 APR 2005 Times Cited: 2
<input type="checkbox"/>	7.	Alba E, Nebro AJ, Troya JM Heterogeneous computing and parallel genetic algorithms JOURNAL OF PARALLEL AND DISTRIBUTED COMPUTING 62 (9): 1362-1385 SEP 2002 Times Cited: 2

Sort by: Times Cited

Analyze Results:

View rankings of the authors, journals, etc. for these records.

Output Records:
 Selected records on page
 All records on page
 Records to
 Bibliographic Fields

 Or add them to the Marked List for later output and more options.

 [0 articles marked]

no!

h=5



Index

Introduction

Quantifying knowledge

The h index

Pros & cons

Practical h

Math facts

➔ Usage

Hands on the h Index: SCOPUS

Using SCOPUS (I)

- Go to <http://www.scopus.com>
- Search for your name and be sure to search “in References”

Search for: in

Scopus: 6,190 Web (0) Patents (52)

Your query: REF(e. alba) [Edit](#) [Save](#) [Save as Alert](#) [RSS](#)

Refine Results [Close](#)

Source Title	Author(s)	Year	Document Type	Subject Area
<input type="checkbox"/> Journal of the American Chemical Society (58)	<input type="checkbox"/> M...	<input type="checkbox"/> 2007 (6)	<input type="checkbox"/> Article (5,200)	<input type="checkbox"/> Agricultural Biological Sciences (1,000)
<input type="checkbox"/> Journal of Agricultural and Food Chemistry (56)	<input checked="" type="checkbox"/> Alba, E. (26)	<input type="checkbox"/> 2006 (860)	<input type="checkbox"/> Short Survey (36)	<input type="checkbox"/> Medicine (1,448)
<input type="checkbox"/> Journal of Chromatography A (44)	<input type="checkbox"/> Asano, N. (21)	<input type="checkbox"/> 2005 (931)		

[Limit to](#) [Exclude](#)

Results: 6,190 Search within results [Go](#)

[Output](#) [Citation tracker](#) [Add to list](#) Select: All Page 1 to 20 [Next](#)

Document (sort by relevance)	Author(s)	Date	Source Title	Cited By
1. <input type="checkbox"/> Long-term effects of drying conditions on the essential oil and color of tarragon leaves during storage Abstract + Refs WebBridge	Arabhosseini, A., Huisman, W., van Bontel, A., Müller, J.	2007	<i>Journal of Food Engineering</i> 79 (2), pp. 561-566	0
2. <input type="checkbox"/> Comparison of volatile components of <i>Marchantia convoluta</i> obtained by supercritical carbon dioxide extraction and petrol ether extraction Abstract + Refs WebBridge	Cao, H., Xiao, J.B., Xu, M.	2007	<i>Journal of Food Composition and Analysis</i> 20 (1), pp. 45-51	0
3. <input type="checkbox"/> Influence of different irrigation strategies in a traditional Comicabra cv. olive orchard on virgin olive oil composition and quality Abstract + Refs WebBridge	Gómez-Rico, A., Salvador, M.D., Moriana, A., Pérez, D., Olmedilla, N., Ribas, F., Frezza, G.	2007	<i>Food Chemistry</i> 100 (2), pp. 568-578	0
4. <input type="checkbox"/> Sulfite determination using sulfite oxidase biosensor based glassy carbon electrode coated with thin mercury film Abstract + Refs WebBridge	Dinckaya, E., Sezgentürk, M.K., Akvilmaz, E., Ertas, F.N.	2007	<i>Food Chemistry</i> 101 (4), pp. 1557-1561	0

that's me!

limit the search to me!



Index

Introduction

Quantifying knowledge

The *h* index

Pros & cons

Practical *h*

Math facts

➔ Usage

Hands on the *h* Index: SCOPUS

Using SCOPUS (II)

- Be sure no one else is in the list and click on “Select All”

Results: 26 Search within results [] Go

Output Citation tracker Add to list Select: All Page 1 to 20 Next ▶

Document (sort by relevance)	Author(s)	Date	Source Title	Cited By
1. <input checked="" type="checkbox"/> Natural language tagging with genetic algorithms Abstract + Refs WebBridge	Alba, E., Luque, G., Araujo, L.	2006	Information Processing Letters 100 (5), pp. 173-182	0
2. <input checked="" type="checkbox"/> Performance of distributed GAs on DNA fragment assembly Abstract + Refs	Alba, E., Luque, G.	2006	Studies in Computational Intelligence 22, pp. 97-115	0
3. <input checked="" type="checkbox"/> Parallel evolutionary multiobjective optimization Abstract + Refs	Luna, F., Nebro, A.J., Alba, E.	2006	Studies in Computational Intelligence 22, pp. 33-56	0
4. <input checked="" type="checkbox"/> Comparative analysis of modern optimization tools for the p-median problem Abstract + Refs WebBridge	Alba, E., Domínguez, E.	2006	Statistics and Computing 16 (3), pp. 251-260	0
5. <input checked="" type="checkbox"/> Hierarchical cellular genetic algorithm Abstract + Refs	Janson, S., Alba, E., Dorronsoro, B., Middendorf, M.	2006	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture	0

- Click on “Citation Tracker”, wait for a while... and that’s it! (or not!?)

The Citation Overview is now being processed

Calculating and Displaying can take a short time, depending on the number of selected documents.

Progress:

If you want to end this process, click Stop.

Stop

Citation Overview Citations received since 1996

Sort documents: year descending Date Range: 2004 to 2006 Go

26 Cited Documents save to list	Citations					
	<2004	2004	2005	2006	subtotal	>2006 total
<input type="checkbox"/> 2006 Natural language tagging with genet...	15	31	31	40	102	0
<input type="checkbox"/> 2006 Parallel evolutionary multiobjectiv...					0	0
<input type="checkbox"/> 2006 Performance of distributed GAs on D...					0	0
<input type="checkbox"/> 2006 Comparative analysis of modern opti...					0	0
<input type="checkbox"/> 2006 Hierarchical cellular genetic algor...					0	0
<input type="checkbox"/> 2006 Computing nine new best-so-far solu...					0	0
Total	15	31	31	40	102	0



Index

Introduction

Quantifying knowledge

The h index

Pros & cons

Practical h

Math facts

➔ Usage

Hands on the h Index: SCOPUS

Using SCOPUS (III)

- Just an important final detail left: sort by “citations descending”!

SCOPUS

Search Sources My Alerts My List My Profile

Quick Search Go

Citation Overview Citations received since 1996

Sort Documents: citations descending (circled) Date Range: 2004 to 2006 Go

		Citations							
		<2004	2004	2005	2006	subtotal	>2006	total	
26 Cited Documents save to list									
<input type="checkbox"/>	Delete	Total	15	31	31	40	102	0	117
1	<input type="checkbox"/> 2002 Parallelism and evolutionary algo...	2	11	18	20	49		51	
2	<input type="checkbox"/> 2001 Analyzing synchronous and asynchron...	6	5	3	2	10		16	
3	<input type="checkbox"/> 2002 Parallel evolutionary algorithms ca...	2	5	2	2	9		11	
4	<input type="checkbox"/> 2002 Improving flexibility and efficienc...	2	2	3	4	9		11	
5	<input type="checkbox"/> 2005 The exploration/exploitation tradeo...			2	7	9		9	
6	<input type="checkbox"/> 2000 Influence of the migration policy i...	3	4	1		5		8	
7	<input type="checkbox"/> 2002 Heterogeneous computing and paralle...		4	1		5		5	
8	<input type="checkbox"/> 2004 The influence of grid shape and asy...			1	1	2		2	
9	<input type="checkbox"/> 2005 Selection intensity in cellular evo...				2	2		2	
10	<input type="checkbox"/> 2004 Parallel LAN/WAN heuristics for opt...				1	1		1	

h=6



Index

Introduction

Quantifying
knowledge

The *h* index

Pros & cons

Practical *h*

Math facts

➡ Usage

Final Comments

- **Computing the *h* index is interesting and indicative**
- **The standards of *h* index in computer science have to be still defined (average/minimum values ...)**
- **Tools that help in computing it exist, but not directly**
- **Tools for searching citations are very important, but at this moment it's mainly an art**
- **Much has to be discussed on how to purge citations for individuals and for groups**
- **The big deal now is to define how to use *h* index**