Re: Best way to hash a set of integers?

Source: http://coding.derkeiler.com/Archive/General/comp.programming/2006−09/msg00603.html

- *From*: Logan Shaw <lshaw−usenet@xxxxxxxxxxxxx>
- *Date*: Wed, 20 Sep 2006 04:49:15 GMT

Digital Puer wrote:

> Ben Pfaff wrote:
>
> > "Digital Puer" <digital_puer@xxxxxxxxxxx> writes:
> >
> > > However, I'm still looking for info on how to best hash a
> > > *set*
> > > of integers;
> >
> > Can you
> > tell us anything about the sets?
>
> I guess I'm looking for a way to compute what could be
> called a multi−dimensional key. Suppose I have three integers
> representing age, height (in cm), and weight (in lbs). I want
> to somehow combine those together into a single key that will be
> my index into a hash table (prior to doing mod hash_table_size).

That's not a set! You could have a person who is 95 years old and 95 lbs. So, "95" can occur twice, so it's not a set. Another reason it's not a set is that the order matters.

So, it's actually a list of integers. Or maybe a record of integers would be a better way to describe it.

Just how to hash these depends on several things. How big will your hash table be? What is the distribution of your numbers? Is it a random or is there some pattern to it? In most of these cases, there will be a pattern. Weight and height will be pretty close to a normal distribution. Age will be different.

You do have one good thing going for you with these kinds of

values: most of the low−order bits will be very evenly distributed. It's extremely unlikely that even ages will greatly outnumber odd ages, or that even weights will greatly outnumber odd ones, etc. So that means that at least the least significant bit in each is close to evenly distributed.

With weight, the 5 least significant bits are likely to be pretty evenly distributed, maybe even 5 or 6 bits. With height, most people will be between 160 cm to 180 c